Lethal Autonomous Weapon Systems Under the Law of Armed Conflict

Homayounnejad, Maziar

Awarding institution:
King's College London

Lethal Autonomous Weapon Systems Under the Law of Armed Conflict

Maziar Homayounnejad

Doctor of Philosophy

2018

King's College London

Abstract

Lethal autonomous weapon systems (LAWS) are essentially weapon systems that, once activated, can select and engage targets without further human intervention. While these are not currently fielded nor *officially* part of any nation's defence strategy, there is ample evidence that many States and defence contractors are currently developing LAWS for future deployment. The main law of armed conflict (LOAC) problem posed by these weapon systems is that lethal action will be taken by machine hardware and control software, rather than human operators exercising deliberative judgment at the point of weapons release.

Specifically, LAWS will follow a set of technical processes, which may operate with super-human accuracy and precision, or with brittleness and potential failure, depending on context and circumstances. In contradistinction, LOAC presupposes sentience and self-awareness, for imposing legal obligations; and human metacognitive judgment, for its effective application in armed conflict. LAWS will possess none of these characteristics in the near-term. Accordingly, such weapon systems may only be lawfully deployed with meaningful human control (MHC), to ensure compliance with the LOAC targeting rules.

The US/NATO Joint Targeting process affords substantial opportunity to ensure MHC in LAWS deployments, because of its highly deliberative planning processes. This will remain true, so long as autonomy does not supplant human judgment in the wider decision-making process, and so long as the (legally-enshrined) 'individual attack' limitation is maintained over and above mere technical viability. Moreover, by regarding precautions in attack as a full LOAC principle, and by integrating more technical personnel into the battle staffs, commanders will be in a stronger position to address the LAWS LOAC challenge. That is, to ensure that *appropriate systems* are deployed in a *suitable operational environment* to undertake *machine-feasible tasks*, along with *appropriate precautionary measures* to sufficiently mitigate civilian risk.

Outside the US/NATO context, it is proposed that the LAWS LOAC challenge be facilitated with the development and publication of an expert LOAC Manual.

Table of Contents

Acknowledgements 8

List of Tables and Figures 9

List of Abbreviations 10

List of Cases and Advisory Opinions 12

List of International Instruments 13

Chapter 1: Introduction 15

1.1 Background and Research Motivation 15

1.2 Drivers of Greater Levels of Autonomy in Weapon Systems 19

1.2.1 Advances in Autonomous Technologies 19

1.2.2 Advantages of Weapons Autonomy 20

1.2.3 Strategic Competition 21

1.3 Methodology and Scope of Thesis 22

1.3.1 Limitations of Methodology and Sources 23

1.3.2 Disciplinary, Factual and Legal Scope of Analysis 26

1.3.2.1 Disciplinary Scope 26

1.3.2.2 Factual Scope 26

1.3.2.3 Legal Scope 26

1.4 Research Question 27

1.5 The Research Puzzle 28

1.6 Thesis and Thesis Structure 29

Chapter 2: Technical Foundations of Military Robotics, Artificial Intelligence and Weapons Autonomy 31

2.1 Introduction 31

2.2 Technical Foundations of Autonomy and Control Systems 32

2.2.1 Landmines as a Simple Model of 'Autonomy' 32

2.2.2 The Instruments of Autonomy and How Autonomy is Achieved 33

2.2.3 Two Essential Corollaries of Autonomy and 'Sense-Think-Act' 36

2.2.3.1 An Essentially Technical Process 36

2.2.3.2 Two Essential Characteristics for Reliable Autonomy 36

2.3 Defining 'Autonomy' in Weapon Systems 38

2.3.1 The Many Faces of 'Autonomy' 38

2.3.2 The Dimensions of 'Weapons Autonomy' 39

2.3.3 Towards a Working Definition of 'Weapons Autonomy' and LAWS 45

2.3.4 Three Clarifying Comments 45

2.3.4.1 The Technical Nature of LAWS 'Discretion' 45

2.3.4.2 Prioritising the Dimensions for Administrability 46

2.3.4.3 The Role of Humans 47

2.4 Weapon Systems Likely to Emerge as LAWS 48

2.4.1 Development of Standalone LAWS 49

2.4.2 Development of Swarms 50

2.4.3 Distant-Future and Existing Autonomy ... 53
2.4.4 A Three-Part Chronology ... 54
2.5 Artificial Intelligence and Automatic Target Recognition ... **56**
2.5.1 Artificial Intelligence ... 56
2.5.1.1 Narrow AI versus *General AI* ... 57
2.5.1.2 Top-Down (Rules-Based) versus *Bottom-Up (Learning) Approaches* ... 57
2.5.1.3 Machine Learning Methods ... 59
2.5.1.4 Modes of Machine Learning ... 61
2.5.2 Deep, Convolutional and Recurrent Neural Networks ... 62
2.5.3 The Inherent Brittleness of AI ... 66
2.5.3.1 Brittleness in Context ... 66
2.5.3.2 Brittleness and the Risk of Learning 'Wrong Lessons' ... 67
2.5.3.3 Transparency, Opacity and the 'Black Box' of Machine Learning ... 68
2.5.4 Automatic Target Recognition ... 71
2.5.4.1 Standard ATR Approaches ... 71
2.5.4.2 Cooperative and Non-Cooperative Targets ... 73
2.5.4.3 Target Indication versus *Target Identification* ... 73
2.5.4.4 GPS Guidance Systems ... 74
2.5.4.5 Vision-Based Guidance Systems ... 75
2.5.5 The Brittle Nature of ATR and Potential Solutions ... 76
2.5.5.1 Sensitivity to Environmental Conditions ... 76
2.5.5.2 Narrow Domains versus *the Broader Context* ... 76
2.5.5.3 Scarcity of Quality Data ... 77
2.5.5.4 Structural ATR Weaknesses and the Proposed TRACE Solution ... 78
2.5.6 Fooling the ATR with Adversarial Examples ... 79
2.5.6.1 The Nature of Adversarial Examples ... 80
2.5.6.2 The Apparent Unavoidability of Adversarial Risk ... 81
2.5.6.3 Can an AI Be Inoculated Against Spoofing? ... 83
2.5.7 Broader Implications for the Military Use of AI ... 84
2.6 Conclusion ... **85**
Chapter 3: Broad Legal Implications of Weapons Autonomy ... **87**
3.1 Introduction ... **87**
3.2 LAWS Will Be Mere Tools, Not Persons ... **87**
3.2.1 The Technical Nature of LAWS ... 89
3.2.2 The Absence of Necessary Human Qualities ... 91
3.2.2.1 LAWS Will Be Neither Sentient nor Self-Aware ... 92
3.2.2.1.1 Sentience ... 92
3.2.2.1.2 Self-Awareness ... 92
3.2.2.1.3 Legal Consequences ... 93
3.2.2.2 Might LAWS Have the Capacity for Metacognition? ... 96
3.2.2.2.1 A Distinctly Human Trait, Necessary for Applying the LOAC ... 97
3.2.2.2.2 Is There an Emerging Machine Metacognition? ... 100
3.2.2.2.3 Machine Metacognition Reconsidered ... 102
3.2.2.2.4 Legal Consequences ... 105
3.2.2.3 The Irrelevance of Human Emotion ... 106
3.2.2.4 Conclusion on the Absence of Human Qualities ... 107
3.3 Implications of Weapons Autonomy for Legal Analysis ... **108**
3.3.1 The Effect of Autonomy on the Assignment, Timing and Character of Operational Decisions ... 109
3.3.1.1 Autonomy Reassigns Operational Decisions ... 110

3.3.1.2 Autonomy Necessitates Earlier Timing of Operational Decisions ... 110
3.3.1.3 Autonomy Changes the Underlying Character of Operational Decisions ... 111
3.3.1.4 Consequence: Autonomy Weakens the Causal Nexus Between Human Decisions, and Specific Actions and Outcomes ... 112
3.3.2 The Machine-Operator Relationship: 'Delegation', Not Abdication ... 112
3.4 Conclusion ... **113**
Chapter 4: 'Meaningful Human Control' in Autonomy ... **115**
4.1 Introduction ... **115**
4.2 The Practical Need for Human Judgment and Control ... **116**
4.2.1 Human Control Over the Machine ... 117
4.2.2 Machine Assistance to Enhance Human Control ... 118
4.2.3 Machine Restrictions on Human Action ... 119
4.3 The Structural Legal Argument for Human Judgment and Control Over an 'Individual Attack' ... **120**
4.3.1 Individual Rules Requiring 'MHC' ... 120
4.3.2 A Broad Standard, Not a Bright-Line Rule ... 121
4.3.3 The Structural Nature of MHC and the Non-Necessity of Additional Rules ... 122
4.4 Human-Machine Interaction 'Touchpoints' ... **124**
4.4.1 Upstream *versus* Downstream Touchpoints ... 126
4.4.2 An Individual, Not a Cumulative Standard ... 127
4.4.3 Core and Derived MHC in Law ... 128
4.5 The Elements of MHC ... **130**
4.5.1 Article 36: MHC Building Blocks in More Detail ... 130
4.5.2 Article 36: Key Elements of MHC ... 132
4.5.3 Horowitz and Scharre: Essential Components of MHC ... 133
4.5.4 ICRC: Distilling MHC From Current-Day Weapon systems ... 134
4.5.5 Common Strands and Elements of MHC ... 135
4.5.6 An Interpretive Aid, Not a Distinct Legal Concept ... 136
4.6 Conclusion ... **140**
Chapter 5: US and NATO Joint Targeting Doctrine as an Expression of Meaningful Human Control ... **142**
5.1 Introduction ... **142**
5.2 Some Key Targeting Definitions and Categories ... **143**
5.2.1 Target ... 143
5.2.2 Targeting and the Targeting Process ... 144
5.2.3 Levels of Warfare and Command ... 145
5.2.4 The Joint Targeting Cycles and Target Categories ... 146
5.2.5 Engagement Categories ... 149
5.2.5.1 Targeted Strike ... 149
5.2.5.2 Tactical-Level Combat ... 150
5.2.5.3 Platform-Defence ... 150
5.3 A Closer Look at the Joint Targeting Cycles ... **151**
5.3.1 The Deliberate Joint Targeting Cycle ... 151
5.3.1.1 Phase 1: (End State and) Commander's Intent, Objectives and Guidance ... 153
5.3.1.2 Phase 2: Target Development (and Prioritisation) ... 153

5.3.1.3 Phase 3: Capabilities Analysis ... 157
5.3.1.4 Phase 4: Commander's Decision, Force Planning and Assignment ... 159
5.3.1.5 Phase 5: Mission Planning and Force Execution ... 160
5.3.1.6 Phase 6: Assessment ... 162
5.3.2 The Dynamic Joint Targeting Cycle ... 162
5.3.2.1 Dynamic Targeting Steps ... 162
5.3.2.2 Inverting the Dynamic Targeting Steps for 'Platform-Defence' ... 164
5.3.3 Preliminary Conclusion on the Joint Targeting Cycles ... 166
5.4 The Central Decision in Targeting ... 167
5.5 Is There Meaningful Human Control Over Attacks Planned Under the Joint Targeting Cycles? ... 169
5.5.1 A Strongly Human Element Within the Joint Targeting Cycles ... 170
5.5.2 A Note on the 'Individual Attack' Limitation ... 172
5.5.3 Is there a Risk of Autonomising the 'Wider Loop'? ... 175
5.5.4 MHC in Weapons Design and Development ... 178
5.6 Conclusion ... 179
Chapter 6: Targeting Law I: How Might LAWS Be Deployed in Compliance with the Principle of Distinction? ... 181
6.1 Introduction ... 181
6.2 The Normative IHL/LOAC Framework ... 182
6.3 A Brief Note on Weapons Law ... 183
6.4 Targeting Law: An Overview ... 186
6.5 The Principle of Distinction ... 187
6.5.1 General Provisions ... 188
6.5.2 Persons ... 191
6.5.2.1 Active Combatants ... 191
6.5.2.1.1 The General Position ... 191
6.5.2.1.2 The Legal Position on Uniforms and Adversarial Examples ... 195
6.5.2.2 Civilians and Other Protected Persons ... 196
6.5.2.3 Civilians Not Protected from Direct Attack ... 199
6.5.2.4 Persons Hors de Combat ... 202
6.5.3 Objects ... 205
6.5.3.1 'Nature' and 'Location' ... 206
6.5.3.2 'Purpose' and 'Use': The Problem of 'Dual-Use' Objects ... 209
6.5.3.3 The 'Definite Military Advantage in the Circumstances Ruling at the Time' ... 215
6.5.3.4 Civilian Objects and Specifically Protected Objects ... 216
6.5.3.4.1 Cultural Property ... 217
6.5.3.4.2 Objects Indispensable for Civilian Survival ... 219
6.5.3.4.3 Infrastructure That May Release Dangerous Forces ... 219
6.5.3.4.4 Medical Capabilities ... 221
6.5.3.4.5 Enhancing Detection by Technical Means ... 222
6.5.3.4.6 Enhancing Confidence in Technical Detection ... 223
6.5.4 Will LAWS be Able to Sense Targeting 'Doubt'? ... 224
6.5.5 Ensuring the Compliance of LAWS Operations with the Principle of Distinction 228
6.6 Conclusion ... 231

Chapter 7: Targeting Law II: Can LAWS Be Deployed in Compliance with the Principle of Proportionality and with Adequate Precautions? **232**

7.1 Introduction **232**

7.2 The Principle of Proportionality **233**

7.2.1 Clarifying Some Pivotal Terms 235

7.2.2 Problematic Compliance with Proportionality in LAWS Deployments 240

7.2.3 Ensuring the Compliance of LAWS Operations with the Principle of Proportionality 245

7.2.3.1 Restricted Deployments 245

7.2.3.2 Thurnher: 'Workarounds' 246

7.2.3.3 Boothby: The Precautionary Value of Process 248

7.2.3.4 Van den Boogaard: Proportionality and the Levels of Warfare 248

7.2.4 Conclusion on Proportionality 252

7.3 Precautionary Measures **252**

7.3.1 The General Obligation Under Article 57(1), AP I 252

7.3.2 Specific Treaty-Based Precautions Under Article 57(2) and (3) 253

7.3.2.1 Target Verification 253

7.3.2.2 Minimise Collateral Damage 256

7.3.2.3 Cancel or Suspend Attacks 258

7.3.2.4 Effective Advance Warning 260

7.3.3 The 'Obligation' to Take Passive Precautions Under Article 58 261

7.3.4 A Potential Interplay Between Articles 57 and 58 262

7.3.5 'Elevating' Precautions to the Status of a LOAC Principle 263

7.3.5.1 Precautions as the Lynchpin of the Targeting Process 263

7.3.5.2 A Precautionary Principle? 265

7.3.6 Potential Applications of a Precautions Principle for LAWS Deployments 267

7.3.6.1 Front-Loading 267

7.3.6.2 Developing Capabilities 268

7.3.6.3 A Precautionary Approach to Online Learning 268

7.3.6.4 Spatio-Temporal Restrictions 269

7.3.6.5 Target Parameters 270

7.3.6.6 Upper Engagement Limits 272

7.3.6.7 Training 275

7.3.6.8 Staffing 276

7.3.7 How to 'Elevate' Precautions to a Full LOAC Principle 277

7.4 Conclusion **279**

Chapter 8: Conclusion **281**

8.1 Linking Back to the Research Question **281**

8.2 Beyond the US/NATO Context: The Value of Standardisation **283**

8.2.1 The Normative Status and Value of a Treaty *versus* a LOAC Manual 285

8.2.2 The Potential Contours of a LOAC Manual on LAWS 287

8.2.3 Potential Challenges to Developing a LOAC Manual on LAWS 291

8.2.4 A Possible Way Forward 292

Bibliography **293**

Acknowledgements

Big thanks to Markus Wagner of Warwick Law School, with whom I spent ten weeks as a visiting scholar. His healthy scepticism of technology and what it can achieve in armed conflict has been particularly formative to the approach I have taken in this project. Equally, Geoffrey Corn of South Texas College of Law Houston has had a significant impact on my thinking. Through both his writings and the time I spent with him as a visiting scholar (and beyond), Geoff's rigorous approach to LOAC firmly encouraged me to apply the law through the lens of the military targeting process. I am grateful for his encouragement in pursuing this particular line of inquiry, which I believe makes for a more realistic analysis. Thanks also to Thomas Schultz for undertaking Law School administration.

There are many others who I have come across, either at King's College or at various conferences and UN meetings, who have been helpful in shaping and often challenging my views for the better. Special thanks to Jürgen Altmann, Ronald Arkin, David Bicknell, Stuart Casey-Maslen, Missy Cummings, Janina Dill, Merel Ekelhof, Ellen Fridland, Tony Gillespie, Ian Goodfellow, James Gow, Mark Gubrud, Kristin Hausler, Ian Henderson, Michael Horowitz, Jasper Hortensius, Matthew Howard, Joshua Hughes, Marissa Kemp, Eliav Lieblich, Jack McDonald, Milton Meza Rivas, Richard Moyes, Wolfgang Richter, Marco Roscini, Paul Scharre and Solon Solomon.

List of Tables and Figures

Table 4.1: Human-machine interaction touchpoints……...……124-126

Figure 2.1: Operation of an anti-personnel mine……...…….32
Figure 2.2: Manually-operated weapon system…….…….34
Figure 2.3: Lethal autonomous weapon system…….…….35
Figure 2.4: Deep neural network and its hidden layers…….…….…….63
Figure 2.5: A RNN's take on breakfast…….……...64
Figure 2.6: Adversarial images……...…….80
Figure 2.7: Adversarial objects…….……...80
Figure 2.8: A simplified two-dimensional manifold…….……...82

Figure 5.1: Targeting categories……147
Figure 5.2: The engagement continuum……151
Figure 5.3: The NATO Joint Targeting Cycle……...152
Figure 5.4: The US Joint Targeting Cycle……..…152
Figures 5.5-5.8: Various targeting activities undertaken by a range of battle staffs ……...172

Figure 6.1: Establishing the 'ground truth' in a relatively complex environment….191

Figure 7.1: The chronology of targeting……264

Figure 8.1: Alternative deployment scenarios and the 'threshold of lawfulness'….282

List of Abbreviations

A2/AD	Anti-access/area-denial
AI	Artificial intelligence
AMW	Air and Missile Warfare (Manual)
AP I	Additional Protocol I
ATR	Automatic Target Recognition
CCM	Convention on Cluster Munitions
CCW	Convention on Certain Conventional Weapons
CDEM	Collateral Damage Estimation Methodology
CIHL	Customary International Humanitarian Law Study
CNN	Convolutional neural network
CODE	Collaborative Operations in Denied Environments
DARPA	Defense Advanced Research Projects Agency
DMA	Definite military advantage
DNN	Deep neural network
DoD	Department of Defense (US)
ECD	Estimated collateral damage
ECMA	Effective contribution to military action
EO/IR	Electro-optical/infrared
F2T2EA	Find, Fix, Track, Target, Engage, Assess
GC	Geneva Convention(s)
GGE	Group of Governmental Experts (meeting)
GPS	Global Positioning System
HVT	High-value target
ICRC	International Committee of the Red Cross
IHL	International humanitarian law
ISR	Intelligence, surveillance and reconnaissance
JFC	Joint Force Commander
JTCB	Joint Targeting Coordination Board
LAWS	Lethal autonomous weapon system(s)
LOAC	Law of armed conflict (Manual)
MAA	Military advantage anticipated
MHC	Meaningful human control

MN-H	Military necessity-humanity (balance)
MoD	Ministry of Defence (UK)
NATO	North Atlantic Treaty Organisation
NGO	Non-governmental organisation
OODA	Observe, Orient, Decide, Act
PGM	Precision-guided munition
PID	Positive identification
PNT	Position, Navigation, Timing
RNN	Recurrent neural network
ROE	Rules of engagement
T&E	Testing and evaluation
TLC	Tactical-level combat
TRACE	Target Recognition and Adaptation in Contested Environments
TS	Targeted strike
TST	Time-sensitive target
UEL	Upper engagement limit
UN	United Nations
V&V	Verification and validation
WO	Weapons operator
XAI	Explainable artificial intelligence

List of Cases and Advisory Opinions

International Court of Justice (ICJ)

Fisheries Case (United Kingdom v. *Norway)* (Judgment) [1951] ICJ Rep 116 (18 December)

North Sea Continental Shelf Cases (Federal Republic of Germany/Denmark; Federal Republic of Germany/Netherlands) (Judgment) [1969] ICJ Rep 3 (20 February)

Delimitation of the Maritime Boundary in the Gulf of Maine Area (Canada/United States of America) [1984] ICJ Rep 246 (12 October)

Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion) [1996] ICJ Rep 226 (8 July)

International Criminal Tribunal for the Former Yugoslavia (ICTY)

Prosecutor v. *Blaškić* (ICTY Trial Judgment) IT-95-14-T (3 March 2000)

Prosecutor v. *Galić* (ICTY Trial Judgment) IT-98-29-T (5 December 2003)

Prosecutor v. *Kordić & Čerkez* (ICTY Appeals Judgment) IT-95-14/2-A (17 December 2004)

Prosecutor v. *Tadić* (ICTY Appeals Judgment) IT-94-1-A (15 July 1999)

International Military Tribunal at Nuremberg

United States v. *List (Wilhelm) et al. (The Hostage Case)* Case No. 7, 19 February 1948 (1950) 11 TWC 1230

Israel

Public Committee against Torture in Israel et al. v. *Government of Israel* et al. (2005) HCJ 769/02 (*Targeted Killings Case*)

List of International Instruments

1907

Annex to Convention (IV) Respecting the Laws and Customs of War on Land: Regulations Concerning the Laws and Customs of War on Land (adopted 18 October 1907, entered into force 26 January 1910) 36 Stat. 2227 TS 539 (Hague Regulations)

1945

Charter of the United Nations (adopted 26 June 1945, entered into force 24 October 1945) 1 UNTS XVI (UN Charter)

Statute of the International Court of Justice (adopted 26 June 1945, entered into force 24 October 1945) 145 BFSP 832 (ICJ Statute)

1949

Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 31 (GC I)

Geneva Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 85 (GC II)

Geneva Convention (III) Relative to the Treatment of Prisoners of War (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 135 (GC III)

Geneva Convention (IV) Relative to the Protection of Civilian Persons in Time of War (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 287 (GC IV)

1954

Convention for the Protection of Cultural Property in the Event of Armed Conflict (adopted 14 May 1954, entered into force 7 August 1954) 249 UNTS 240 (Hague Convention)

1969

Vienna Convention on the Law of Treaties (adopted 23 May 1969, entered into force 27 January 1980) 1155 UNTS, 331 (Vienna Convention)

1977

Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3 (AP I)

1980

Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects (adopted 10 October 1980, entered into force 2 December 1983, amended on 21 December 2001) 1342 UNTS 137 (CCW)

Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons (adopted 10 October 1980, entered into force 2 December 1983) 1342 UNTS 171 (Protocol III to the CCW)

1993

Annex I to Protocol I Additional to the Geneva Conventions of 1949: Regulations Concerning Identification (as amended on 30 November 1993, entered into force 1 March 1994) (Amended Annex I)

1994

Convention on the Safety of United Nations and Associated Personnel (adopted 9 December 1994, entered into force 15 January 1999) 2051 UNTS 363

1996

Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices (adopted 10 October 1980, amended 3 May 1996, entered into force 3 December 1998) 2048 UNTS 93 (Amended Protocol II to the CCW)

1997

Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction (adopted 18 September 1997, entered into force 1 March 1999) 2056 UNTS 241 (Mine Ban Treaty)

1998

Rome Statute of the International Criminal Court (adopted 17 July 1998, entered into force 1 July 2002) 2187 UNTS 3; UN Doc. A/CONF.183.9 (Rome Statute)

2003

Protocol on Explosive Remnants of War (adopted 28 November 2003, entered into force 12 November 2006) 2399 UNTS 100 (Protocol V to the CCW)

2008

Convention on Cluster Munitions (adopted 30 May 2008, entered into force 1 August 2010) 2688 UNTS 190 (CCM)

Chapter 1

Introduction

1.1 Background and Research Motivation

Lethal autonomous weapon systems (LAWS) are essentially "weapon system[s] that, once activated, can *select* and *engage* targets without further intervention by a human operator".[1] While they are not yet fielded,[2] LAWS are currently in early stage development in numerous States,[3] and may well be deployed within the next decade or so. For the United States (US) in particular, broader military autonomy is an integral part of its Third Offset Strategy,[4] which aims to bolster US conventional deterrence in the face of declining force structures *vis-à-vis* near-peer adversaries.[5] More recently, the 2018 *National Defense Strategy* has pledged to "invest broadly in military applications of autonomy, artificial intelligence, and machine learning" as part of its goal of 'modernising key capabilities', to gain competitive military advantage.[6] Accordingly, the development, fielding and deployment of LAWS is arguably likely to occur in the near-term, as advancing technologies combine with a perceived sense of military necessity and strategic competition.

LAWS will differ from more traditional means of warfare in ways that will give rise to specific legal challenges under the law of armed conflict (LOAC), also known as international humanitarian law (IHL).[7] For the purpose of this thesis, it is significant

---

[1] US Department of Defense (DoD), *Directive No. 3000.09: Autonomy in Weapon Systems* (21 November 2012, incorporating *Change 1*, 8 May 2017) (hereafter, *Directive 3000.09*), 13 <http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf> accessed 30 September 2018.

[2] But see the Israeli *Harpy*, in 2.4.

[3] See Maziar Homayounnejad, 'Regulating Lethal Autonomous Weapon Systems I: Assessing the Sense and Scope of 'Autonomy' in Emerging Military Weapon Systems', *TLI Think! Paper 76/2017* (2017), 22-23 <https://ssrn.com/abstract=3027540> accessed 30 September 2018.

[4] See Pentagon Memorandum, 'The Defense Innovation Initiative' (15 November 2014) <http://archive.defense.gov/pubs/OSD013411-14.pdf> accessed 30 September 2018.

[5] Cheryl Pellerin, 'Deputy Secretary: Third Offset Strategy Bolsters America's Military Deterrence', *US Department of Defense News* (31 October 2016) <https://www.defense.gov/News/Article/Article/991434/deputy-secretary-third-offset-strategy-bolsters-americas-military-deterrence/> accessed 30 September 2018.

[6] US DoD, *Summary of the 2018 National Defense Strategy of the United States of America* (DoD, January 2018) (hereafter, *2018 NDS*), 7.

[7] These two terms will be used interchangeably in this thesis, though with a preference for LOAC.

that the human operator is kept 'out-of-the-loop' in the critical functions,[8] thereby leaving sensory equipment and software algorithms to determine the nature, extent and timing of lethal action against specific targets. This is in contradistinction to manned and remotely-piloted systems, where lethal action is taken by human operators exercising deliberative reasoning at the point of weapons release, and it raises questions on the capacity of LAWS to be deployed in compliance with the LOAC targeting rules. However, such legal problems will not necessarily include an accountability gap, so long as commanders and their staffs retain meaningful human control over LAWS.[9] At a minimum, control will be exercised over broader (strategic and operational) parameters during the targeting process;[10] hence, commanders should be held to account if such parameters are set negligently or recklessly, and if this leads to unlawful actions on the battlefield.

Rather, the main LOAC-based problem concerns decision-making under relatively more abstract circumstances. Deliberative human decisions on lethal targeting will have to be made in *earlier phases* of the targeting cycle, at *locations further removed* from the intended strike site, and with *less accurate knowledge of concrete threats or specific civilian risks*.[11] Thus, with software algorithms that lack the cognitive sophistication of a human mind, there will need to be an effective 'division of labour' between man and machine, to optimise the allocation of tasks in accordance with their respective cognitive strengths.[12]

Consider the following broad examples.

- Where LAWS are deployed in ostensibly benign battlefields, yet systems begin to struggle to distinguish between protected and non-protected entities.[13]
- An abundance of civilians unexpectedly appears, leading to a heightened risk of 'disproportionate' attack; especially considering the indeterminate and context-specific nature of proportionality, which a LAWS will not be able to assess in fully autonomous mode.[14]

---

[8] That is, to select (find, fix, track) and engage (target, engage, assess) a target. See 2.3.2 and 5.3.2.1.

[9] See Chapter 4.

[10] See Chapter 5.

[11] See Chapter 2, especially 2.3.4.3.

[12] See 4.2.

[13] See Chapter 6.

[14] See 7.2.

- Since earlier and more geographically removed targeting decisions are made in relatively more abstract circumstances, commanders may miss opportunities to implement effective and LAWS-relevant precautionary measures. In turn, this may lead to insufficient civilian risk mitigation once the concrete battlefield situation materialises.[15]

All of these scenarios may lead to potentially unlawful (machine) actions on the battlefield; certainly, they increase civilian risk. Underlying the concern is the fact that systems are inherently brittle and they cannot always adapt to changing circumstances;[16] they do not possess agency, and they cannot apply the law.[17] Only responsible humans are the addressees of LOAC, hence the associated obligations under targeting law are *exclusively* for commanders, their battle staffs and weapons operators to meet. Accordingly, the LAWS IHL/LOAC challenge is for human decision-makers to ensure that *appropriate systems* are deployed in a *suitable operational environment* to undertake *machine-feasible tasks*, along with *appropriate precautionary measures* to sufficiently mitigate civilian risk.

An alternative response to the above legal challenge is to institute a comprehensive and pre-emptive ban on the development, testing, production, deployment and use of LAWS.[18] Indeed, since May 2013, this is exactly what is being pursued at the United Nations *Convention on Certain Conventional Weapons* (*CCW*), by a coalition of non-governmental organisations (NGOs) known as the *Campaign to Stop Killer Robots*.[19] As of 13 April 2018, 23 States have also joined the call for a pre-emptive ban,[20] thus providing a sizeable minority of governmental support. Meanwhile, in July 2015 the *Future of Life Institute* (*FLI*) published an open letter – now signed by almost 4,000 artificial intelligence (AI) researchers – calling for "a ban on offensive autonomous

---

[15] See 7.3.

[16] See Chapter 2, especially 2.5.5-2.5.7.

[17] See Chapters 2 and 3.

[18] Human Rights Watch, *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch, 2012), 46-47. See also the Berlin Statement (2010) and the Original and 2014 Mission Statements of the *International Committee for Robot Arms Control* <http://icrac.net/statements/> both accessed 4 October 2018.

[19] See <https://www.stopkillerrobots.org/> accessed 10 May 2018.

[20] Campaign to Stop Killer Robots, *Retaining Human Control of Weapons Systems* (9-13 April 2018) <https://www.stopkillerrobots.org/wp-content/uploads/2018/03/KRC_Briefing_CCWApr2018.pdf> accessed 4 October 2018 (listing 22 States, after which China curiously joined the call for a ban).

weapons beyond meaningful human control".[21] The widespread media attention this received ensured that concern about 'killer robots' has, to a certain extent, mobilised the general public. More recently (August 2017), the *FLI* issued a second open letter, this time not explicitly calling for a ban, but listing the apparent dangers of lethal autonomy and imploring State Parties at the *CCW*'s first Group of Governmental Experts (GGE) meeting to "find a way to protect us all from these dangers".[22]

However, while ban proponents have vigorously argued their position,[23] there remain near-insurmountable practical issues that would make both the negotiation/signing and enforcement of a ban treaty unlikely to succeed.[24] Chief amongst these are the potential military advantage and force multiplying capabilities of LAWS, which will very likely defeat consensus towards a ban;[25] as well as numerous difficulties with verifying compliance, in the event that a ban is formally agreed.[26] In addition, there are strong arguments pointing to the normative desirability of allowing potentially beneficial nascent technologies to develop.[27] Chief amongst these are the opportunity to remove from the battlefield the human frailties and imperfections[28] that often lead to erroneous

---

[21] FLI, 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers' (28 July 2015) <https://futureoflife.org/open-letter-autonomous-weapons/> accessed 4 October 2018.

[22] FLI, 'An Open Letter to the United Nations Convention on Certain Conventional Weapons' (21 August 2017) <https://futureoflife.org/autonomous-weapons-open-letter-2017> accessed 4 October 2018.

[23] For example, Human Rights Watch (n 18); Human Rights Watch, *Precedent for Preemption: The Ban on Blinding Lasers as a Model for a Killer Robots Prohibition* (Human Rights Watch, November 2015); Human Rights Watch, *Making the Case: The Dangers of Killer Robots and the Need for a Preemptive Ban* (Human Rights Watch, 2016); Thompson Chengeta, 'Measuring Autonomous Weapon Systems Against International Humanitarian Law Rules' (2016) 5 Journal of Law and Cyber Warfare 63; Peter Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making' (2012) 94 International Review of the Red Cross 687; Noel E. Sharkey, 'The Evitability of Autonomous Robot Warfare' (2012) 94 International Review of the Red Cross 787.

[24] Rebecca Crootof, 'Why the Prohibition on Permanently Blinding Lasers is Poor Precedent for a Ban on Autonomous Weapon Systems', *Lawfare* (24 November 2015) <https://www.lawfareblog.com/why-prohibition-permanently-blinding-lasers-poor-precedent-ban-autonomous-weapon-systems> accessed 4 October 2018.

[25] Ibid.; Sean Watts, 'Autonomous Weapons: Regulation Tolerant or Regulation Resistant?' (2016) 30 Temple International & Comparative Law Journal 177, 187 (pointing to a potential "Balkanized sector of weapons law").

[26] Rebecca Crootof, 'The Killer Robots are Here: Legal and Policy Implications' (2015) 36 Cardozo Law Review 1837.

[27] Charles J. Dunlap, Jr, 'To Ban New Weapons or Regulate Their Use?' *Just Security* (3 April 2015) <https://www.justsecurity.org/21766/guest-post-ban-weapons-regulate-use/> accessed 4 October 2018.

[28] For example, hunger, tiredness, hatred, fear, the instinct for revenge and a wilful disrespect for IHL/LOAC.

targeting, or even war crimes;[29] and to exploit more accurate, precise and responsive technology that could lower civilian casualties.[30]

For these reasons, the following thesis will assume that LAWS will not be banned; instead, they will very likely be developed, fielded and deployed by States, but also specifically regulated,[31] or otherwise be the subject of a non-binding LOAC Manual.[32] The preference is for the latter, as this may be quicker to conclude and – being written by a group of LOAC experts – is likely to attract greater respect from the world's militaries, which have to apply the rules in the field.[33]

1.2 Drivers of Greater Levels of Autonomy in Weapon Systems

While not central to the legal arguments, it is worth briefly reflecting on the reasons why weapons autonomy is becoming a reality; hence, why the application of the LOAC targeting rules to LAWS will soon be inevitable. These drivers of greater autonomy can be categorised into three groups.

1.2.1 Advances in Autonomous Technologies

First, there are substantial advances in AI, which are making reliable autonomy increasingly viable.[34] In turn, these developments are driven by a "perfect storm of [cheap] parallel computation, bigger data, and deeper algorithms",[35] which are revolutionising both civilian and military technologies. For example, AI systems can now beat humans at poker[36] and Go,[37] and they can even beat a real Top Gun pilot in

---

[29] Ronald C. Arkin, 'Lethal Autonomous Systems and the Plight of the Non-Combatant' (2013) 137 AISB Quarterly 4; Marco Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified' (2014) 90 International Law Studies 308.

[30] Ibid.; Michael N. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) Harvard National Security Journal Features 1.

[31] Crootof (n 26).

[32] Steven Groves, 'A Manual Adapting the Law of Armed Conflict to Lethal Autonomous Weapons Systems', *Margaret Thatcher Center for Freedom, Special Report No. 183*, (The Heritage Foundation, 7 April 2016) <http://thf-reports.s3.amazonaws.com/2016/SR183.pdf> accessed 4 October 2018.

[33] See 8.2, on the merits of producing a LOAC Manual on LAWS.

[34] See Chapter 2.

[35] Kevin Kelly, 'The Three Breakthroughs That Have Finally Unleashed AI on the World', *Wired* (27 October 2014) <http://www.wired.com/2014/10/future-of-artificial-intelligence/> accessed 4 October 2018.

[36] Cade Metz, 'Inside Libratus, The Poker AI That Out-Bluffed the Best Humans, *Wired* (1 February 2017) <https://www.wired.com/2017/02/libratus/> accessed 4 October 2018.

[37] Dan Silver et al., 'Mastering the Game of Go Without Human Knowledge' (2017) 550 Nature 354.

a simulated aerial dogfight.[38] With such rapidly-advancing technologies offering greater speed, accuracy and precision, it is arguably only a matter of time before they are integrated into critical weapons functions.

1.2.2 Advantages of Weapons Autonomy[39]

Second, LAWS will offer a number of distinct (military) advantages that will give deploying forces an upper hand. These include the following.

- The ability to **operate without communications links**. This offers certain *tactical* benefits, such as minimising the risk of hacking and jamming in the field.[40] It also offers major *strategic* benefits, which are of growing importance with the recent return to Great Power Competition.[41] For example, near-peer adversaries like Russia and China have stepped up their 'grey zone' activities,[42] and have erected numerous anti-access/area-denial (A2/AD) networks,[43] which create a communications-denied environment on their shores. This prevents US/NATO forces from having the capability to 'look deep and strike deep' into their territory, thereby limiting the projection of Western power into these distant theatres.[44] Accordingly, the *capability* to strike in communications-denied environments is now of premium strategic value to US/NATO forces, even if such capability never actually results in a kinetic attack.[45]
- The **speed and responsiveness of machine action** relative to human performance. This arises because of electronic data-processing speeds, compared with the human neuromuscular delay of around 0.25 seconds. It

---

[38] Nick Ernest and Kelly Cohen, 'Genetic Fuzzy Based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions' (2015) 6 Journal of Defense Management 139.

[39] For fuller analyses of these advantages, see Maziar Homayounnejad, 'The Lawful Use of Autonomous Weapon Systems for Targeted Strikes (Part 1): Concepts, Advantages and Technologies', *TLI Think! Paper 11/2018* (2018), 16-26 <https://ssrn.com/abstract=3158170> accessed 4 October 2018.

[40] Jeff Hecht, 'Did Iran Capture US Drone by Hacking its GPS Signal?' *New Scientist* (16 December 2011).

[41] *2018 NDS* (n 6).

[42] 'Pride and Prejudice: The Odds on a Conflict Between the Great Powers', *The Economist Special Report: The Future of War* (27 January 2018) (referring to cyber-attacks, assassination, fake news, propaganda, bribery, and military intimidation).

[43] Ibid.; General Sir Nicholas Carter, 'Dynamic Security Threats and the British Army', *Speech Delivered to the Royal United Services Institute* (22 January 2018) <https://rusi.org/event/dynamic-security-threats-and-british-army> accessed 4 October 2018.

[44] Interview with former US Deputy Defense Secretary, Robert O. Work, in Octavian Manea, 'The Role of Offset Strategies in Restoring Conventional Deterrence', *Small Wars Journal* (4 January 2018) http://smallwarsjournal.com/jrnl/art/role-offset-strategies-restoring-conventional-deterrence> accessed 4 October 2018.

[45] Homayounnejad (n 39), 20.

manifests itself in military advantages (responding quicker to enemy fire); humanitarian benefits (cancelling attacks quicker upon detecting an apparently unlawful situation); and preserving combat capability (taking more evasive defensive manoeuvres when coming under enemy fire).[46]

- **Cost and resource utilisation**. With no crews required to operate individual units (save for ground crews and a weapons operator per *n* tactical units), LAWS will clearly result in major cost-savings. Moreover, with autonomous pattern-recognition capabilities, a LAWS can loiter for extended periods and can simply alert a human operator when there is a relevant situational change that may call for kinetic action. This obviates the need to always have military personnel standing watch, and it affords an opportunity to reallocate manpower to other tasks.[47]

1.2.3 Strategic Competition[48]

Finally, while no State has *officially* declared a policy of integrating LAWS into its armed forces, certain official policies and statements have pointed to a competitive element that may *ultimately* drive the fielding of LAWS. For example, in December 2017 China announced its National AI Strategy, part of which aims to promote a military-civil fusion in which the People's Liberation Army (PLA) seeks to capitalise on the disruptive military potential of AI.[49] The aim is to take the PLA from today's 'informatized' way of warfare towards 'intelligentized' warfare.[50] In September 2017, Russian President Vladimir Putin stated that "the one who becomes the leader in [the AI] sphere will be the ruler of the world",[51] and this approach was clearly reflected in Russia's opposition to a LAWS ban at the November 2017 GGE meeting.[52] Arguably,

---

[46] Ibid., 20-25.

[47] Ibid., 25-26.

[48] See Michael C. Horowitz et al., Strategic Competition in an Era of Artificial Intelligence', *CNAS Series on AI and International Security* (July 2018) <https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf?mtime=20180716122000> accessed 4 October 2018.

[49] Ibid., 13.

[50] Elsa B. Kania, 'China is On a Whole-of-Nation Push for AI. The US Must Match It', *Defense One* (8 December 2017) <https://www.defenseone.com/ideas/2017/12/us-china-artificial-intelligence/144414/> accessed 4 October 2018.

[51] Associated Press, 'Putin: Leader in Artificial Intelligence Will Rule World', *CNBC* (4 September 2017) <https://www.cnbc.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html> accessed 4 October 2018.

[52] Patrick Tucker, 'Russia to the United Nations: Don't Try to Stop Us From Building Killer Robots', *Defense One* (21 November 2017) <https://www.defenseone.com/technology/2017/11/russia-united-nations-dont-try-stop-us-building-killer-robots/142734/> accessed 4 October 2018.

this position is not new, given the 2015 statement of Chief of General Staff Valery Gerasimov's that in the near-future, Russia may have "a fully roboticized unit…capable of independently conducting military operations".[53] For their part, US officials have not been as publicly bold, though former Deputy Secretary of Defense Robert Work has posed the question:

> If our competitors go to Terminators…and it turns out the Terminators are able to make decisions faster, even if they're bad, how would we respond?[54]

Other US officials have dubbed this "The Terminator Conundrum",[55] and it essentially points to a growing offence-defence dynamic. Namely, while US policy in *Directive 3000.09* currently requires a man-in-the-loop in every kill decision, this may well come under pressure if near-peer adversaries begin to field and deploy LAWS.

1.3 Methodology and Scope of Thesis

With the above in mind, this thesis will examine the application of the LOAC targeting rules of international armed conflict to the likely eventual deployment and use of LAWS. Primarily (though not exclusively), the focus will be on the 1977 Additional Protocol I[56] (AP I), cross-referenced with customary international law. Formally, the difference between the two is that only the 174 States party to AP I are bound by it, whereas all States are bound by customary law. In practice, much of AP I is customary and, therefore, binds all States; where this is not the case, the thesis will comment on the difference, which will be relevant to AP I non-Party States such as the US. In addition, extensive interpretive guidance will be gleaned from the *AP I Commentary*[57] and the *Commentary* to the *Air and Missile Warfare Manual*.[58]

---

[53] See Robert O. Work, 'Deputy Secretary of Defense Speech', *CNAS Defense Forum* (14 December 2015) <http://www.defense.gov/News/Speeches/Speech-View/Article/634214/cnas-defense-forum> accessed 4 October 2018.

[54] Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (Norton, 2018), 8.

[55] Colin Clark, The Terminator Conundrum: VCJCS Selva on Thinking Weapons', *Breaking Defense* (21 January 2016) <https://breakingdefense.com/2016/01/the-terminator-conundrum-vcjcs-selva-on-thinking-weapons/> accessed 4 October 2018.

[56] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3.

[57] Yves Sandoz, Christophe Swinarski and Bruno Zimmermann (eds.), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Convention of 12th August 1949* (Martinus Nijhoff, 1987).

[58] Program on Humanitarian Policy & Conflict Research at Harvard University, *Commentary on the HPCR Manual on International Law Applicable to Air and Missile Warfare* (v2.1, Harvard College, 2010) (hereafter, *AMW Manual Commentary*).

1.3.1 Limitations of Methodology and Sources

It should be emphasised that neither the *AP I Commentary* nor the *AMW Manual* or its *Commentary* are sources of law, hence they do not have normative status. The authoritative approach to determining the sources of international law is set out in Article 38(1) of the Statute of the International Court of Justice.[59] This recognises three distinct sources.

- **Treaty law**,[60] which is defined as "an international agreement concluded between States in written form and governed by international law".[61] The focus is on "rules expressly recognized by…States", irrespective of the particular name or designation of the document.[62]
- **Customary law**,[63] which arises when there is sufficient evidence of State practice (by a critical mass of affected States over a sufficient period of time) coupled with *opinio juris*; that is, a belief on the part of the State that such practice is required by a rule of international law.[64]
- **General principles of law**,[65] which are broad legal concepts derived from national legal systems. Hence, they are so fundamental that they transcend national boundaries, and they may be used to fill gaps in the law.[66]

The above reflects the fact that only States can make international law, not commentators or experts acting in their personal capacity. Thus, within this thesis only AP I and other cited treaties enjoy normative status (see below on customary law).

---

[59] See Article 38(1)(a)-(c), Statute of the International Court of Justice (adopted 26 June 1945, entered into force 24 October 1945) 145 BFSP 832 (hereafter, ICJ Statute).

[60] Article 38(1)(a), ICJ Statute.

[61] Article 2(1)(a), Vienna Convention on the Law of Treaties (adopted 23 May 1969, entered into force 27 January 1980) 1155 UNTS, 331.

[62] Article 38(1)(a), ICJ Statute. Hence, titles such as 'annex', 'protocol', 'convention', or 'statute' – all of which will be encountered in this thesis – are merely terms, which do not affect the normative status of the document.

[63] Article 38(1)(b), ICJ Statute.

[64] *North Sea Continental Shelf Cases (Federal Republic of Germany/Denmark; Federal Republic of Germany/Netherlands)* (Judgment) [1969] ICJ Rep 3, ¶ 77.

[65] Article 38(1)(c), ICJ Statute.

[66] Lord Lloyd-Jones (Justice of the UK Supreme Court), 'General Principles of Law in International Law and Common Law', *Speech Delivered to the Conseil d'Etat, Paris* (16 February 2018) <https://www.supremecourt.uk/docs/speech-180216.pdf> accessed 1 April 2019.

On the other hand, Article 38(1) also recognises "the teachings of the most highly qualified publicists…as *subsidiary means* for the *determination* of rules of law".[67] Namely, the scholarly writings of experts may act as learned viewpoints from which the substance of the treaty and customary law rules may be extracted. Arguably, this is also where treaty commentaries and LOAC Manuals (and their commentaries) fit in.[68] The former constitutes major references for the interpretation and application of treaties, and they are often written by experts acting in their personal capacity.[69] The latter are restatements of the law that reflect the consensus of large groups of experts, also acting in their personal capacity.[70] None of these emanate from State entities, much less do they have binding force, though they are highly influential as 'subsidiary' means to 'determine' the substance and application of the law.[71] As noted above, they will provide interpretive guidance for applying treaty and/or customary law.

A more controversial document, which will inform some of the analysis in Chapter 6, is the *Interpretive Guidance on the Notion of Direct Participation in Hostilities* issued by the International Committee of the Red Cross (ICRC).[72] The project behind it started with a substantial group of experts, but as consensus did not emerge on key matters, the ICRC issued the *Guidance* as an institutional document.[73] Importantly, the *Guidance* seeks to provide recommendations concerning the interpretation of LOAC in a specific context,[74] thus its legal status is arguably like that of a LOAC Manual. However, the absence of expert consensus and the abundance of scholarly criticism (see 6.5.2.3) will undoubtedly affect its persuasiveness, which can be expected to be lower and more variable than that of a LOAC Manual or its commentary.

---

[67] Article 38(1)(d), ICJ Statute (emphasis added).

[68] William H. Boothby, *Conflict Law: The Influence of New Weapons Technology, Human Rights and Emerging Actors* (TMC Asser Press, 2014), 85-87.

[69] See, for example, Foreword to the *AP I Commentary*, xiii (stating "the ICRC…allowed the authors their academic freedom, considering the Commentary above all as scholarly work, and not as a work intended to disseminate the views of the ICRC").

[70] Boothby (n 68), 86.

[71] For example, the *AMW Manual Commentary* notes, at 3, that the *Manual* itself "does not have binding force, but hopefully it will serve as a valuable resource for armed forces in the development of rules of engagement, the writing of domestic military manuals, the preparation of training courses and – above all – the actual conduct of armed forces in combat operations".

[72] ICRC, *Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law* (ICRC, 2009).

[73] Boothby (n 68), 78.

[74] ICRC (n 72), 9.

An important point should be made concerning customary international law. This is one of the recognised sources of international law,[75] but independently researching it is beyond the scope of this thesis. The most rigorous approach would be to find real-world evidence of State practice and, separately, of corresponding *opinio juris* from which to determine specific norms. Alternatively, *opinio juris* "can be tested by induction based on an analysis of a sufficiently extensive and convincing [State] practice".[76] This, however, is not the focus of the thesis; instead, the ICRC's *Customary International Humanitarian Law* Study[77] will be cited and cross-referenced with treaty law. The ICRC Study is arguably one of the most extensively researched and relevant restatements of the customary law of armed conflict, for which it provides valuable *evidence* of custom.[78] Yet, it is not an actual normative source of custom and, being a largely scholarly endeavour, its legal status is like that of a LOAC Manual.[79] Thus, while the Study is a useful *aid* to determining customary law, it is not definitive, and its methodology, rules, and statements are all open to criticism by commentators.[80]

Finally, a document that will be cited in parts of Chapters 3, 6 and 7 is the national military manual. These are prepared by individuals employed for that purpose – usually within a State defence bureaucracy – and they will tend to express the view of the relevant State on the rules of international law that bind it, and on how those rules should be interpreted.[81] Thus, national military manuals represent the *opinio juris* and, potentially, the State practice of the issuing State. In that sense, the manuals do not in themselves have any normative status, though they may contribute to the formation of customary international law. For the purpose of this thesis, they will also provide interpretive insight, albeit as restatements of the law issued by an individual State.

---

[75] Article 38(1)(b), ICJ Statute.

[76] *Delimitation of the Maritime Boundary in the Gulf of Maine Area (Canada/United States of America)* [1984] ICJ Rep 246, ¶ 111. In practice, however, the ICJ has used a mixture of induction, deduction, and assertion, with an apparently greater reliance on the latter. See Stefan Talmon, 'International Law: The ICJ's Methodology Between Induction, Deduction and Assertion' (2015) 26 The European Journal of International Law 417.

[77] Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law, Vol. 1: Rules* (CUP, 2005).

[78] Boothby (n 68), 76-77.

[79] Namely, in Article 38(1), ICJ Statute, the Study is recognised under Sub-Paragraph (d), not (b).

[80] See, for example, Letter from John Bellinger III, Legal Adviser, US Dept. of State and William J. Haynes, General Counsel, US Dept. of Defense, to Dr Jakob Kellenberger, President, International Committee of the Red Cross, Regarding Customary International Law Study [3 November 2006] (2007) 46 ILM 514. For broader commentary, see Elizabeth Wilmshurst and Susan Breau (eds.), *Perspectives on the ICRC Study on Customary International Humanitarian Law* (CUP, 2011).

[81] Boothby (n 68), 65-66.

1.3.2 Disciplinary, Factual and Legal Scope of Analysis

1.3.2.1 Disciplinary Scope

As noted in 1.1, the technology of LAWS will raise some novel issues affecting legal compliance, and this arguably calls for a strongly interdisciplinary analysis. Accordingly, this thesis will draw on substantial insights from computer science, robotics, psychology, philosophy, human-machine interaction, and military targeting doctrine. These non-legal contributions will serve to contextualise the analysis of the black-letter rules, and to evaluate more effectively how LOAC compliance can be secured in the light of the new technologies. This is to be distinguished from 'international law and [subject X]' methods,[82] which are not the focus of this thesis. The latter often displace the black-letter approach by proposing different ways of perceiving the law, hence they are arguably more suited to policy and *lex ferenda* arguments.[83] By contrast, this thesis is concerned with an accurate application of the *lex lata*, albeit in a specific technological and military procedural context.

1.3.2.2 Factual Scope

As the central thesis of this study concerns the formal targeting process as a conduit for meaningful human control (MHC), the legal analysis will be applied in the context of US and North Atlantic Treaty Organisation (NATO) targeting doctrine. This is justified on the basis that many of the States within NATO are expected to be major LAWS developers and/or users, and because their elaborate targeting cycles arguably provide a model of best practice for other States to emulate.[84] As will be argued in the thesis, the US/NATO targeting process offers particular value in ensuring the safe and lawful deployment of LAWS.

1.3.2.3 Legal Scope

The scope of the legal analysis itself is confined to the conduct of hostilities in international armed conflict (IAC), and this raises three immediate limitations. First,

---

[82] See Andrea Bianchi, *International Law Theories: An Inquiry into Different Ways of Thinking* (OUP, 2017) (outlining 14 different approaches to international law, including positivism, the New Haven School, critical legal studies, international relations theory, feminist jurisprudence, and the law and economics perspective, amongst others).

[83] See, for example, Stephen Ratner and Ann-Marie Slaughter, 'Appraising Methods of International Law: A Prospectus for Readers' (1999) 93 American Journal of International Law 291 (using a fictional LOAC problem to reveal different modes of analysis and different outcomes from seven of the international law approaches).

[84] Geoffrey S. Corn, 'War, Law, and the Oft Overlooked Value of Process as a Precautionary Measure' (2015) 42 Pepperdine Law Review 419.

there will be no focus on the *jus ad bellum* which, while potentially important, raises a very different set of legal and practical concerns.[85] Second, as the LOAC rules are assumed to be the *lex specialis* in an IAC, there will be no analysis of international human rights law, much less the complex interplay between that and LOAC. Third, the thesis will not apply the LOAC rules governing non-international armed conflict (NIAC) either. The latter two exclusions are justified by the likely limitations of near-term LAWS technologies, which will demand the relatively simpler and more machine-perceptible environment of an IAC.[86] In any event, the return to Great Power Competition noted in 1.2.2, above, makes it likely that there will be sufficient military utility from, hence demand for, LAWS that are designed for IAC deployments.

Two further limitations should be noted. First, as commanders can safely assume all weapons in their arsenal are lawful *per se*, subject to prescribed restrictions on use, there will be no examination of weapons law, save for a brief note in Chapter 6. Second, as the legal analysis will assume the 'good faith commander', there will be no consideration of international criminal law, except for a brief mention in Chapter 7, purely for interpretive purposes.

All information contained in the thesis is correct as at 1 May 2018, shortly following the second GGE meeting in April. Where possible, however, a few more recent developments have been included.

1.4 Research Question

This thesis will seek to address the following research question: "To what extent can US/NATO forces apply the existing LOAC targeting rules, to ensure the lawful deployment and use of near-term LAWS?"

In doing so, the thesis will also examine the extent to which machines may be expected to operate reliably on decisions that *would* be taken by a human operator, in a counterfactual scenario of manned or remotely-piloted targeting; bearing in mind, however, that any shortfalls will likely be addressed within the Joint Targeting process.

---

[85] See Paul Scharre, 'Autonomous Weapons and Operational Risk', *CNAS Ethical Autonomy Project* (February 2016) <http://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf> accessed 22 September 2018 (discussing unanticipated interactions between adversarial systems during peacetime, and the risk of triggering a 'flash war').

[86] See Chapter 2, especially 2.4.4 on the 'crawl-walk-run' approach to LAWS development.

1.5 The Research Puzzle

Currently, a great deal of the LAWS IHL/LOAC literature applies a selective sample of the legal rules, often with vastly different assumptions on the nature of technology. Consequently, pro- and anti-LAWS positions tend to become more entrenched. As Scharre points out:

> Some envision autonomous weapons as more reliable and precise than humans, the next logical evolution of precision-guided weapons, leading to more humane wars with fewer civilian casualties. Others envision calamity, with rogue robot death machines killing multitudes.[87]

The former position is taken by Arkin,[88] Schmitt[89] and Sassóli,[90] amongst others;[91] while the latter position is taken by the ban proponents.[92] However, Scharre continues:

> It is entirely possible that both [visions] come true, with autonomous weapons making war more precise and humane *when they function properly*, but causing mass lethality *when they fail*.[93]

For the positive outcome to materialise, it is arguably necessary for LAWS-deploying forces to understand both capacities and limitations and, crucially, to avoid anthropomorphising LAWS; lest they put too much faith in brittle technologies that are prone to failure, when poorly deployed.[94] Indeed, this point was emphasised by the third GGE meeting in August 2018, which included vigilance against anthropomorphisms as one of its guiding principles.[95]

---

[87] Scharre (n 54), 347.

[88] Arkin (n 29).

[89] Schmitt (n 30).

[90] Sassóli (n 29).

[91] For example, Kenneth Anderson, Daniel Reisner and Matthew C. Waxman, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems' (2014) 90 International Law Studies 386.

[92] See the various works cited in n 23.

[93] Scharre (n 54), 347 (emphasis added).

[94] See Chapter 3, especially 3.2.

[95] *Report of the 2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems* (31 August 2018) UN Doc. CCW/GGE.2/2018/3, ¶ 26(h) <https://www.unog.ch/80256EDD006B8954/(httpAssets)/20092911F6495FA7C125830E003F9A5B/$file/2018_GGE+LAWS_Final+Report.pdf> accessed 4 October 2018 ("In crafting potential policy measures, emerging technologies in the area of [LAWS] should not be anthropomorphized").

Accordingly, the original contribution made by the following thesis is two-fold.

(1) To address the anthropomorphism concern with an in-depth review of relevant technologies, how they differ from human qualities, and an application of these to broad LOAC targeting requirements.[96]

(2) An in-depth application of the LOAC targeting rules to LAWS, with a greater level of detail than seen in other works, and with a relatively more considered application to technological and operational targeting realities. This will offer a more representative and contextual analysis of the potential lawfulness of LAWS deployments.[97]

1.6 Thesis and Thesis Structure

The thesis of this study is that the existing LOAC targeting rules are likely to be sufficient for regulating LAWS deployments by US/NATO forces; assuming the rules are applied within a formal targeting process, in which tasks that require deliberative thinking can be undertaken by humans before or during deployment.

The thesis is divided into six substantive chapters.

Chapter 2 will provide a detailed introduction to the technical foundations of military robotics and AI. It will define weapons autonomy and LAWS, and it will explain what kinds of weapon systems are likely to emerge as 'autonomous' in the near-term. Crucially, this chapter also demonstrate that LAWS will follow a set of *technical* processes, which may operate with super-human accuracy and precision (in circumstances for which they were designed); or with brittleness and potential failure (in unexpected situations, or in circumstances that fall outside their design envelope).

Chapter 3 will build on this, to examine the broad legal implications of weapons autonomy. Here, it will be argued that only accountable humans can have legal obligations and, more importantly, the prevalence of standards/principles over rules in LOAC means that machines cannot 'apply' the law of war either. That said, it is possible for machines to execute technical processes that commanders will have anticipated in their legal assessment of a planned attack, and this will form the outer boundaries of lawful weapons autonomy.

---

[96] Chapters 2-4.

[97] Chapters 5-7.

Chapter 4 examines the MHC concept, and argues that not only is this a practical necessity for the safe operation of a weapon system, but it is also a structural *legal* requirement in relation to an 'individual attack'. Equally important, MHC is not restricted to the point of weapons release, but occurs at seven 'touchpoints', both upstream (far before deployment) and downstream (closer to, or during deployment). Crucially, this chapter also details the elements of MHC and argues that – to preserve the military necessity-humanity balance – the concept is best regarded as an interpretive aid, rather than a distinct legal concept.

Chapter 5 provides an in-depth examination of the US/NATO targeting process, and it draws a practical distinction between targeted strikes (against specific and unique targets) and tactical-level combat (against broader target sets). Importantly, this chapter demonstrates how the targeting process distributes human decision-making across six distinct phases and provides a basis for MHC over individual attacks. Accordingly, it may be appropriate to reconceptualise the 'critical functions' from the narrow technical phenomenon of target *recognition*, to the broader human-led targeting *process*.

With all the foregoing analyses in mind, Chapters 6 and 7 apply the LOAC targeting rules to potential LAWS deployments, focusing on the principles of distinction, proportionality and precautions in attack. Here, it will be seen that there are ample opportunities to deploy and use autonomous weapons in compliance with the individual LOAC rules, so long as there is an effective 'division of labour' between man and machine, in accordance with their respective cognitive strengths. Concretely, commanders will have to ensure that a) appropriate systems are deployed, b) in a suitable operational environment, c) to undertake machine-feasible tasks, along with d) appropriate precautionary measures to sufficiently mitigate civilian risk.

Chapter 8 concludes and provides some final thoughts, along with a brief consideration of two regulatory options for LAWS: negotiating a treaty or drafting a LOAC Manual.

Chapter 2
Technical Foundations of Military Robotics, Artificial Intelligence and Weapons Autonomy

2.1 Introduction

The following chapter contains four main sections, which will lay down some technical concepts, definitions and assumptions that are necessary for a more valid legal analysis. First, 2.2 explains the foundations of autonomy and the control systems that are an integral part of military robotics. Here, it will be seen that when undertaking autonomous attack, LAWS will execute highly sophisticated, albeit *technical* processes that preclude these devices from having any agency. Hence, they cannot 'apply legal rules' as such, and it will be necessary for humans to constrain their actions. Second, 2.3 examines the various dimensions of autonomy, and it develops a working definition of 'lethal autonomous weapon systems' (LAWS). As will be seen, autonomy in weapon systems narrowly concerns stochastic behaviour in the *critical functions* of selecting and engaging targets, and this necessarily has a humanitarian impact. Other forms of autonomy *may* have humanitarian effects, but not necessarily or directly so. With this in mind, 2.4 delineates the kinds of weapon systems likely to emerge as 'autonomous' in the near-term. These will mostly be *wide-area search-and-attack* platforms and munitions, deployed at sea, in the air, and on land, respectively. Initially, they will focus on anti-material targeting in uncluttered environments, but will gradually take on broader attack missions in a 'crawl-walk-run' approach. Finally, 2.5 provides a detailed account of the most important technologies necessary for autonomous attack: artificial intelligence and automatic target recognition. This section will demonstrate that while LAWS are expected to outperform humans in many narrow tasks, such as object recognition, the systems will remain *brittle* and unable to adapt when the environment or context for action changes. Accordingly, it will be vital to keep a human somewhere in the loop, to ensure that as machine autonomy is harnessed, it is also kept in check by deliberative human control, which can employ 'common sense' to adapt to the situation at hand. More concerning is the 'black box' nature of the complex software that will control many LAWS, yet will behave in inexplicable and counterintuitive ways that will be open to exploitation by the adversary. The implications will be explored in subsequent chapters.

2.2 Technical Foundations of Autonomy and Control Systems

Any assessment of the legal consequences of weapons autonomy must be based on a clear understanding of what this concept entails: how it is achieved, and its main characteristics and guises.[1] The in-depth technical analysis is less important than the impact that autonomy has on the assignment, timing and character of operational decisions; and on the interactions between a LAWS, its operators/supervisors and those subject to its effects.[2] The following will therefore sketch an overview of the technical aspects of machine autonomy, as a primer for the analysis of legal implications in Chapter 3 and human control in Chapter 4.

2.2.1 Landmines as a Simple Model of 'Autonomy'

While LAWS will be vastly more sophisticated than landmines, the two share an essential characteristic that arguably justifies (cautious) comparison: lack of contemporaneous human control at the point of 'trigger pull'.

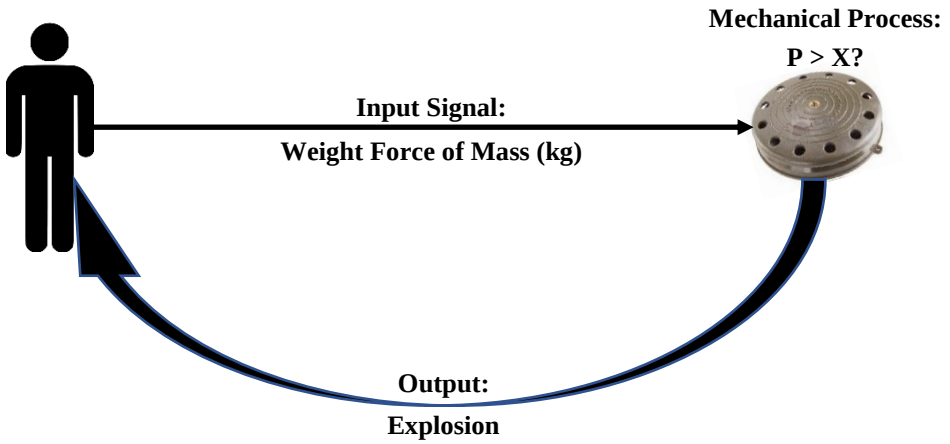

Figure 2.1: Operation of an anti-personnel mine. Source: Adapted from Moyes (n 3).

Moyes summarises the 'selection' and engagement process of a landmine, as follows.[3]

[1] Tim McFarland, 'Factors Shaping the Legal Implications of Increasingly Autonomous Military Systems', (2015) 900 International Review of the Red Cross 1313.

[2] Ibid.

[3] Richard Moyes, 'Autonomous Weapons Systems Policy', *Talk Delivered for MIT Course 6.S099: Artificial General Intelligence* (17 April 2018) <https://www.youtube.com/watch?v=U6lJI-NSfBY&t=2847s> accessed 8 May 2018.

- A person steps on the mine, providing an **input signal** consisting of a 'weight force of mass' in kilograms.
- Pressure plates within the mine are pressed down (P). A basic **mechanical algorithm** consisting of a minimum pressure threshold (X) operates to determine a response.
- Should P > X, the mine delivers an **output** by way of an explosion, which kills or seriously injures the person providing the input signal.

Fundamental to an understanding of LAWS is that the weapon will operate a simple loop, being victim-activated with no other person intervening.[4] Moreover, the entire process is mechanical (input→algorithm→output) and offers no *immediate* input of human reasoning or judgment, despite this being necessary when a civilian weighing enough to trigger P > X walks into the mined area. That said, precautionary human judgment can be involved in deciding *where* to lay the mines; for *how long* before demining the area; and in *deliberate design features*, for example, in setting an appropriate X value, or integrating features that cause the mines to self-destruct or self-deactivate.[5]

To be sure, landmines are 'automatic' rather than 'autonomous' weapons, hence the above model may be overly simplistic.[6] Nonetheless, it provides a useful entry point for understanding the challenges that may arise with LAWS deployments, as well as a justification to seek insights and ideas from the landmines regime.[7]

2.2.2 The Instruments of Autonomy and How Autonomy is Achieved

Modern control theory "deals with the behavior of dynamical systems[8] with inputs, and how their behavior is modified by feedback".[9] In this (interdisciplinary) branch of engineering and mathematics, autonomous systems are regarded as a *control system* which consists of two separate components:[10]

---

[4] Ibid.

[5] See 7.3.6.1 on front-loading.

[6] See 2.3 on the levels of autonomy.

[7] See 7.3.6.4-7.3.6.5 on spatio-temporal restrictions and target parameters.

[8] That is, a system or process exhibiting internally dynamic behaviour. See S. Simrock 'Control Theory' in Daniel Brandt (ed.), *CAS CERN Accelerator School* (CERN, 2008), 73 <https://cds.cern.ch/record/1100534/files/p73.pdf> accessed 8 May 2018.

[9] Laurence Buen Suarez, *Control Theory Fundamentals* (Delve, 2017), Preface.

[10] Ibid. See also Zdzislaw Bubnicki, *Modern Control Theory* (Springer, 2005).

- the *plant*, which is the system or process to be controlled; and
- the *controller*, which is a device consisting of both hardware and software or, more precisely, an *executor* and a *control algorithm*.

In a LAWS context, the *control system* is the overall weapon system. The *plant* might be a drone or a gun turret – equipment which, if not autonomous, would be controlled by a human operator. The *controller* would be the combination of hardware and software that manages this equipment, in accordance with parameters programmed by a developer.[11] Before elaborating on how this may operate, first consider how a manually-operated weapon system might work.

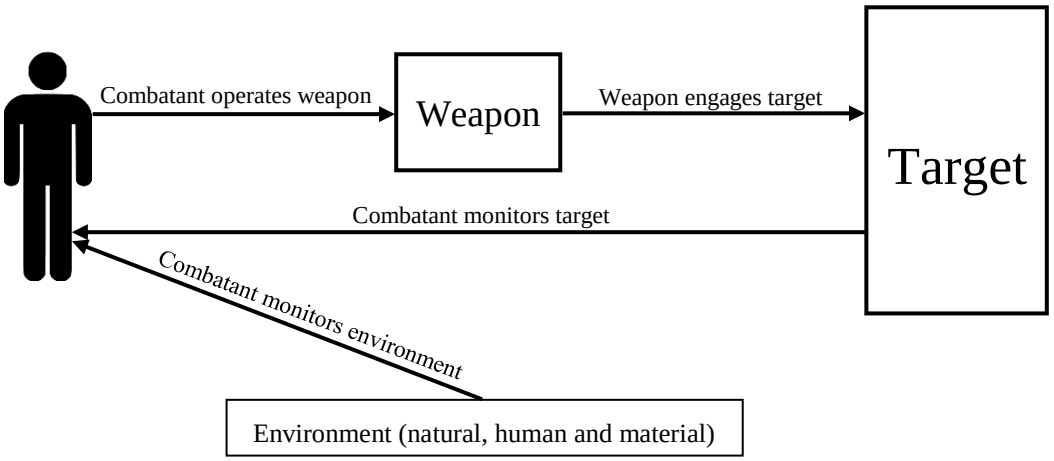

Figure 2.2: Manually-operated weapon system. Source: adapted from Figure 1, McFarland (n 1).

In Figure 2.2, above, the solid arrows show typical interactions between the various components. The combatant/weapons operator (WO) monitors both the target and the operational environment; and he operates the weapon, which only engages the target after direct human input. Accordingly, the entire decision-making cycle is conducted by the individual WO.[12]

In air combat, pilots often discuss Boyd's 'observe, orient, decide, act' (OODA) loop as the cognitive process they undertake when engaging enemy aircraft.[13] By 'getting inside' the enemy's OODA loop – reacting while he is still trying to assess the situation

[11] McFarland (n 1), 1318.

[12] Subject to mission Rules of Engagement.

[13] Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (Norton, 2018), 23.

– a fighter pilot forces the enemy to restart their own OODA loop and remain in a state of perpetual "confusion and disorder". This causes the enemy "to over and under react to activity that appears simultaneously menacing as well as ambiguous, chaotic, or misleading."[14] Importantly, where a manned weapon system is used, this OODA loop remains largely the (recurring) decision-making cycle of the human WO, as above.

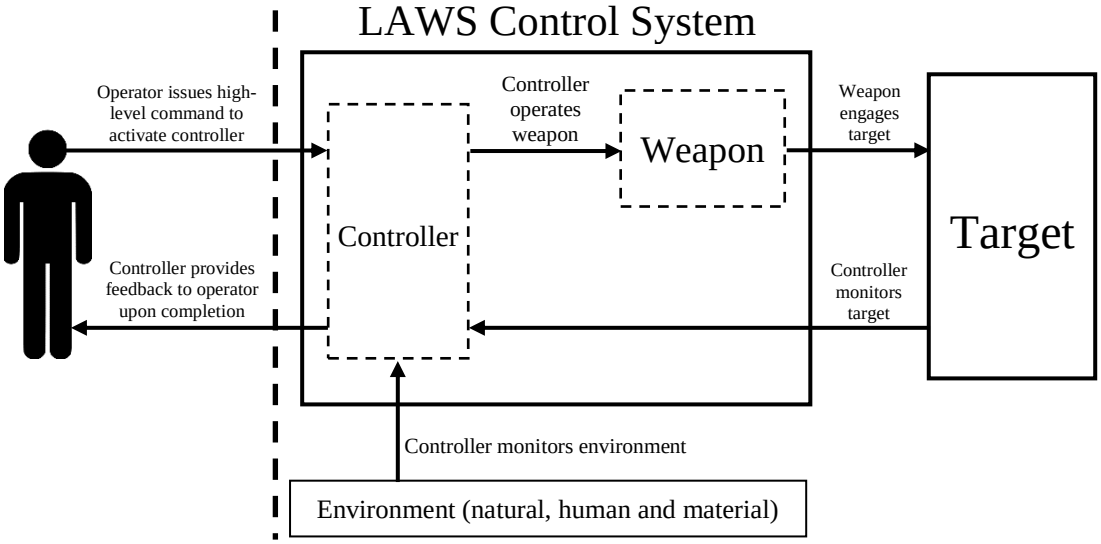

Figure 2.3: Lethal autonomous weapon system. Source: adapted from Figure 2, McFarland (n 1).

By contrast, Figure 2.3 conceptually outlines how a control system may be expected to operate in a LAWS context. The human WO activates the weapon system and issues high-level commands; much like a commander issuing orders to ground troops, but without the agency or capacity for moral reasoning often seen in the latter. The WO's/system developer's understanding of how to control the weapon is expressed in software codes programmed into the controller.[15] To some extent, the software controller then "steps into the shoes" of the WO, operating the weapon system as well as monitoring the target, the environment and the weapon system itself, via a series of distinctly technical processes.[16] Accordingly, the system takes over a large part of the human combatant's OODA loop,[17] with a speed advantage that may compress this into micro or nanoseconds, thus moving OODA towards becoming a "perceive and act"

[14] John R. Boyd, 'Patterns of Conflict' in Chet Richards and Chuck Spinney (eds.), *Defense and the National Interest* (January 2007), 7 <http://www.dnipogo.org/boyd/patterns_ppt.pdf> accessed 8 May 2018.

[15] McFarland (n 1), 1324.

[16] Ibid.

[17] See 3.3 on the impact of reassigning the OODA loop on the legal analysis of operational decisions.

vector.[18] This it does via the 'sense-think-act' paradigm, in which the robot proceeds in three consecutive, yet rapid steps.[19]

- **Sense**: perceives its environment, using a range of sensory hardware and software.[20]
- **Think**: processes the raw data in accordance with its software and programming.[21]
- **Act**: interacts with and disturbs its environment, potentially taking kinetic action.[22]

2.2.3 Two Essential Corollaries of Autonomy and 'Sense-Think-Act'

2.2.3.1 An Essentially Technical Process

Much of the above clearly demonstrates a series of technical processes that rely on statistical correlations. In particular, sensing software, such as those tasked with computer vision, undertake *pattern recognition* via pixel-by-pixel analysis of raw data and a comparison with image patterns stored in memory.[23] There is little or no equivalence with human sensory perception and processing, and the more complex the software artefact, the more unpredictable is the output. As will be seen in 2.5.6, this may lead to inexplicable and counterintuitive errors, which – in a safety-critical context – could have catastrophic consequences.

2.2.3.2 Two Essential Characteristics for Reliable Autonomy

Another consequence of following a technical process is that any **tasks** assigned to a machine must meet certain minimum criteria based on *precision* (how well-defined the task is, and whether it can be specified in programmable rules) and *tangibility* (how quantifiable the expected outcomes are).[24] Two further factors affecting the complexity or viability of the task are: *dimensionality*, in terms of requiring a single action to complete the task, or numerous sequential decisions and actions; and the level

---

[18] US Air Force, *Unmanned Aircraft Systems Flight Plan, 2009-2047* (USAF HQ, 18 May 2009), 41.

[19] See, for example, Vincent Boulanin and Maaike Verbruggen, *Mapping the Development of Autonomy in Weapon Systems* (SIPRI, November 2017), 8-11.

[20] For example, video camera and computer vision; radar, infrared and sonar sensors; and GPS.

[21] See 2.5.1-2.5.2.

[22] This occurs using actuators (robotic 'muscles'), which power the robot's effectors (physical devices that manipulate the environment). Examples of actuators include electric motors and hydraulic cylinders, while effectors include wheels, legs, wings, grippers, or weapons.

[23] Boulanin and Verbruggen (n 19), 8-9.

[24] Ibid., 13.

of *interaction* with other autonomous agents, which may become difficult to model, as (especially human) behaviour can be unpredictable.[25]

The **environment** also affects the viability of achieving autonomy. Important considerations include whether the environment is *fully observable* or only *partially observable* through sensors; *structured* or *unstructured*; *uncluttered* or *cluttered*; *static* or *dynamic*; *deterministic* or *stochastic*; *cooperative* or *adversarial*.[26] In each pairing, the first adjective tends towards a simple environment, which is more easily modelled in advance and amenable to autonomous action; the second tends towards a more complex environment, which poses challenges for autonomous attack. The only constant in warfare is its adversarial nature. Beyond that, both attack tasks and operational environments can be very diverse, posing a range of challenges of varying degrees of difficulty for compliance with the law of armed conflict (LOAC), or international humanitarian law (IHL).

This overview of military robotics suggests two opposing realities. While LAWS will undoubtedly be sophisticated machines, able to undertake a range of warfighting roles with super-human senses, speed, accuracy and precision, they will nonetheless be mechanical and electronic devices that merely apply technical processes. This latter point strikes a similarity with the operation of landmines, and it underscores two realities: one factual, the other legal. *Factually*, LAWS acting alone will likely remain brittle and will lack the sophistication of deliberative human judgment.[27] As Cummings points out, "The best operating conditions for [LAWS] are those that promote a high-fidelity world model with low environment uncertainty". [28] This will be further illustrated in 2.5.4-2.5.5, in the context of automatic target recognition. *Legally*, the technical nature of a LAWS underscores the lack of agency in these systems, though it does affect the assignment and character of operational decisions, as well as the machine-operator relationship, thereby reallocating some legal responsibilities. The broader legal effects of this point will be revisited in Chapter 3.

---

[25] Ibid.

[26] Ibid. These are not all mutually exclusive pairings, and some of them will tend to overlap.

[27] This point on brittleness – technical proficiency in narrow domains but weakness in understanding the broader context – will be further illustrated in 2.5.3, 2.5.5 and 2.5.6.

[28] ML. Cummings, 'Artificial Intelligence and the Future of Warfare', *Chatham House Research Paper* (January 2017), 4 <https://www.chathamhouse.org/publication/artificial-intelligence-and-future-warfare> accessed 10 May 2018.

2.3 Defining 'Autonomy' in Weapon Systems

Arriving at a satisfactory definition of 'weapons autonomy' and 'LAWS' is more than merely an academic exercise. A **legal** definition will have jurisdictional consequences, as it will determine precisely *which* weapon systems are subject to a LAWS regulation treaty or a LOAC Manual; and whether we are discussing future systems only, or are looking to include existing ones too.[29] A **working** definition, which is the focus here, will set the boundaries of the current analysis. In that regard, this thesis will define weapons autonomy as a concept that applies only to near-future weapon systems. Indeed, given the purpose of LAWS legal and policy discussions is to bring potentially 'problematic' weapon systems under a specific governance regime, it would arguably serve little practical use – and needlessly expend much analytical and regulatory effort – to catch existing systems that operate well without any specifically applicable regulation.[30] Likewise, to define LAWS in a way that only catches hypothetical and distant-future systems would be to miss the point of regulating weapons for humanitarian impact.[31]

2.3.1 The Many Faces of 'Autonomy'

In its purest form, 'autonomy' is a broad concept, deriving from the Greek 'auto' (self) and 'nomos' (law) to mean *self-ruling* or *self-governing*.[32] This has different meanings in different disciplines, with at least five possible variants: political, philosophical, legal, moral[33] and technical.[34] In a LAWS context, it is only the *technical* sense of the word that matters, at least as a starting point. However, even 'technical autonomy' is not defined in any single way, and it is possible to discern two broad but interrelated

[29] Michael C. Horowitz, 'Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems' (2016) 30 Temple International & Comparative Law Journal 85.

[30] Paul Scharre and Michael C. Horowitz, 'An Introduction to Autonomy in Weapon Systems', *CNAS Ethical Autonomy Series Working Paper* (February 2015) <https://s3.amazonaws.com/files.cnas.org/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf> accessed 10 May 2018 (arguing, at 17, that it may even be inimical to the aims of IHL if an overly broad definition is adopted, which prohibits or restricts the use of precision-guided munitions (PGMs) that tend to *reduce* civilian casualties in war).

[31] See, for example, UK Ministry of Defence, *Joint Doctrine Publication 0-30.2: Unmanned Aircraft Systems* (Development, Concepts and Doctrine Centre, August 2017), 13 (defining 'autonomous system' as one that "is capable of *understanding higher level intent* and direction") (emphasis added).

[32] Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate, 2009), 43.

[33] D. Gracia, 'The Many Faces of Autonomy' (2012) 33 Theoretical Medicine and Bioethics 57.

[34] 'Technical autonomy' is specific to robotics and the computer sciences.

groups of definitions here:[35] the first focuses on the *human-machine relationship*;[36] the second focuses on the system's *internal manipulation of its own capabilities*.[37] As will be seen below, 'weapons autonomy' incorporates both of these dimensions, but must also integrate at least one more aspect (the task being performed) to account for the specific military context, and the humanitarian focus of the rules that will govern the development, deployment and use of these systems.

2.3.2 The Dimensions of 'Weapons Autonomy'

In line with the above, Scharre and Horowitz consider that there are three distinct 'dimensions of autonomy',[38] which this thesis will use to frame a working definition of LAWS. First, there is the **level of human control**, which concerns the relationship between the machine and its human operator, as well as task allocation between the two. This also finds reflection in the seminal Human Rights Watch report, *Losing Humanity*,[39] which distinguished between three categories.[40]

- 'Human-*in*-the-Loop Weapons', where the human both selects and engages the target, albeit via the medium of the machine.
- 'Human-*on*-the-Loop Weapons', where the machine selects and engages the target, but under human oversight and with the option of manual override.
- 'Human-*out-of*-the-Loop Weapons', where the machine selects and engages the target *without* any further human interaction.[41]

The concept of autonomy gravitates towards humans being *out-of*-the-loop, although, as will be seen below, lower forms of autonomy may well be associated with the other categories.[42] Furthermore, while the 'level of human control' accords with some of the broader definitions of technical autonomy alluded to above, *weapons* autonomy is a much narrower and more specific term. Namely, to fully capture the essence of

---

[35] Henry Hexmoor et al., 'A Prospectus on Agent Autonomy' in Henry Hexmoor et al. (eds.), *Agent Autonomy* (Springer, 2003), 3-4.

[36] See, for example, Michael A. Goodrich and Alan C. Schultz, 'Human-Robot Interaction: A Survey' (2007) 1 Foundations and Trends in Human-Robot Interaction 203.

[37] See, for example, Defense Science Board, US Department of Defense (DoD), *The Role of Autonomy in DoD Systems* (Office of the Under Secretary of Defense for ATL, July 2012), 1.

[38] Scharre and Horowitz (n 30), 6-7.

[39] Human Rights Watch, *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch, 2012).

[40] Ibid., 2.

[41] Ibid. Note that human-*out-of*-the-loop systems still require initial activation by a human.

[42] See (note and text accompanying) n 139.

machine autonomy in a weapons context, it is necessary to qualify the 'level of human control' with internal system features and a sense of lethality.

Accordingly, a second dimension that should also be considered is **machine complexity**. This refers to the 'intelligence' of the system, and is also often divided into three broad categories.[43]

- *Automatic* implies that the system exhibits a simple, mechanical response to environmental input, such as how a basic thermostat works. In a military context, a landmine or trip-wire would be considered 'automatic', as these involve a direct linear relationship between cause (triggering of a pressure threshold) and effect (explosion), with no real decision-making process in between.
- *Automated* weapons take several environmental inputs into a decision-making process, which is relatively complex, yet grounded in a predictable rules-based system.[44] These typically operate on IF-THEN-ELSE functions, to determine a response, similar to a computer spreadsheet. In a military context, air and missile defence systems are considered 'automated'; or, more precisely, they can be *set to operate* in automated mode in relation to selecting and engaging specific targets.[45]
- *Autonomous* systems are those that "execute some kind of self-direction, self-learning *or* emergent behavior that is not directly predictable from an inspection of its code".[46]

Wagner points out that, in contradistinction to the first two categories, autonomous systems will exhibit two unique features.[47] First, they will have the capacity to make

[43] Scharre and Horowitz (n 30), 6.

[44] See (notes and text accompanying) nn 170-174.

[45] Alternatively, these systems can be set to operate with a human-in-the-loop in these critical functions. If so, the system reverts to being 'automatic', in that a human-launched missile (e.g. from the *Patriot* missile battery) senses the speed and trajectory of its target and automatically alters its flight path to keep in line with the target. This is similarly true of a system like the *Phalanx*, which automatically adapts the direction of its Gatling gun, but only to keep its stream of ammunition aiming towards the human-selected target.

[46] Scharre and Horowitz (n 30), 6 (emphasis added to highlight that these are alternatives; namely, machine learning is an option, but not a requirement to meet the 'threshold' for an autonomous system).

[47] Markus Wagner, 'Autonomous Weapon Systems', *Max Planck Encyclopedia of Public International Law* (2016) <http://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e2134> accessed 10 May 2018.

"discretionary decisions";[48] hence, they will be able to "react, independently, to a changing set of circumstances without necessitating the interference of a human operator".[49] Second, autonomous systems will require no human input on which *specific target* to select and engage; with which *specific munition* to engage that target; or on the *timing of weapons release*.[50] Accordingly, LAWS will be 'goal driven' systems that undertake complex decision-making processes and – in complex operational environments – will be able to react to concrete situations that neither their designers nor deploying commanders may have anticipated.

The above classification also finds reflection in the ICRC's report,[51] which distinguished between 'remotely controlled' systems (tele-operated devices with some automatic features, to support human control by the pilot), 'automated' and 'autonomous' systems;[52] the latter two approximating more closely to Scharre and Horowitz's categories of the same designation.

Finally, there is the **task to be performed by the machine**. This is crucial for determining whether or not machine autonomy can be used in accordance with the law, as different actions will carry different levels of risk if systems malfunction, or are poorly designed or deployed.[53] For example, both a mechanical thermostat and an anti-tank mine have humans 'out-of-the-loop'; but if the sensors become over-sensitive, the latter may kill civilians, whereas the former will be a mere inconvenience. More recently, there are the autonomous take-off and carrier landing capabilities of the US *X-47B* stealth drone, hailed as a technical breakthrough when they were successfully demonstrated in 2013.[54] Under the 'machine complexity' dimension, these may be considered 'autonomous'. However, neither take-off nor landing involve any lethal

[48] Ibid., ¶ 6. On the meaning of LAWS 'discretion', see (notes and text accompanying) nn 77-82.

[49] Ibid.

[50] Ibid.; Markus Wagner, 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems' (2014) 47 Vanderbilt Journal of Transnational Law 1371, 1383.

[51] International Committee of the Red Cross (ICRC), *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects* (Expert Meeting of 26-28 March 2014) (ICRC, November 2014), 62.

[52] Ibid., 62 and 64.

[53] Scharre and Horowitz (n 30); Cummings (n 28).

[54] Brandon Vinson, 'X-47B Makes First Arrested Landing at Sea', *Navy News Services* (10 July 2013) <http://www.navy.mil/submit/display.asp?story_id=75298> accessed 10 May 2018 (emphasising the difficulty of carrier-based landings and the unprecedented nature of autonomous integrated carrier operations).

engagement or damage to property, and both usually occur far away from any actual or potential contact zone; alone, they offer no warfighting capability and pose no perils for civilians or other protected persons or objects. Hence, they are unlikely to qualify as 'weapons autonomy' for IHL/LOAC purposes. By contrast, the critical functions of *selecting* and *engaging* targets for attack – including the secondary critical tasks of *selecting the munition* and the *timing of weapons release* – all have very real humanitarian impacts.[55]

To illustrate the point, it is instructive to consider the work of the US Department of Defense (DoD) Working Group on Autonomous Weapon Systems.[56] Prior to drafting *Directive 3000.09*,[57] the Working Group analysed a selection of existing (mostly supervised or semi-autonomous) technologies and case studies of past catastrophic errors, as well as measures that could have prevented them.[58] These included the following.

- The wrongful shooting down of Iran Air flight 655 by the *Aegis Combat System* (on board the USS *Vincennes*) in 1988, which killed all 290 passengers and crew members on board.[59]
- Two fratricide ('friendly fire') incidents in 2003 – one against a Royal Air Force *Tornado GR4* jet and another hitting a US Navy *F/A 18 Hornet* jet. Both incidents involved the *MIM-104 Patriot* missile battery, and they killed both pilots and the *Tornado* navigator.[60]

---

[55] For example, selecting the wrong target may violate the principle of *distinction*. Releasing a munition with an unnecessarily large blast radius, or releasing it while civilian heat signatures are still visibly dispersing, may violate the *proportionality* principle; almost certainly the obligation to take feasible *precautions in attack*. See Chapters 6 and 7 for further.

[56] This was established by the Under-Secretary of Defense for Policy, to formulate the official US definition of 'autonomous weapon system' and, more generally, to draft US policy on LAWS by way of *Directive 3000.09*.

[57] US DoD, *Directive No. 3000.09: Autonomy in Weapon Systems* (21 November 2012, incorporating *Change 1*, 8 May 2017) <http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf> accessed 10 May 2018.

[58] Colonel Richard Jackson, 'Autonomous Weaponry and Armed Conflict', *ASIL Panel Discussion* (10 April 2014) <https://www.youtube.com/watch?v=duq3DtFJtWg> accessed 10 May 2018.

[59] Gene I. Rochlin, 'Iran Air Flight 655 and the USS *Vincennes*: Complex, Large-Scale Military Systems and the Failure of Control' in Todd R. La Porte (ed.), *Social Responses to Large Technical Systems: Control or Anticipation* (Springer Science + Business Media, 1991) (pointing to 'scenario fulfilment', whereby personnel under pressure carry out standard training scenario responses, and ignore contradictory information that should lead to a different outcome).

[60] See John K. Hawley, 'Patriot Wars: Automation and the Patriot Air and Missile Defense System', *CNAS Ethical Autonomy Series* (January 2017) <https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-

In each case, it was found that it was in the *selection* and *engagement* of the target – namely, those functions most closely related to lethality – where human judgement and the LOAC both applied, yet the potential for catastrophe was the greatest; and indeed where fatal mistakes were actually made, but with no clear lines of accountability.[61] Accordingly, to maintain the humanitarian focus of IHL/LOAC, the relevant tasks that should warrant a legal or (self)-regulatory response are exclusively the (primary and/or secondary) *critical functions* discussed above.[62] Again, this approach is strongly reflected in the ICRC report, which itself distils this focus on the critical functions from a number of other definitions of weapons autonomy.[63]

Furthermore, as Scharre and Horowitz point out, autonomy may be a broad concept in that it can apply to any function *within* a weapon system but, as mentioned above, 'autonomous *weapon'* is an inherently narrower term and can *only* refer to the critical functions.[64] Here, the authors are acknowledging that the whole purpose of a weapon system is to *attack* a target, be that in offence or defence. Hence, to autonomise non-critical functions, while the selection and engagement of targets remain manually operated, should preclude that autonomy – however technically advanced – from making the system an 'autonomous *weapon*' as such. Autonomous weapons are necessarily autonomous in the critical functions. That said, as will be seen in Chapter 5, in the case of US and NATO forces the 'critical functions' should arguably be reconceptualised from the narrow technical phenomenon of target *recognition*, to the broader human-led targeting *process*. This is both to take into account the vastly more deliberative process of target selection, with its meaningful checks and balances; as well as the creeping autonomy in areas like intelligence and target development, which may undermine this.

FINAL.pdf> accessed 10 May 2018 (detailing the causes of the fratricides as a complex mix of failures, of which machine autonomy was only one).

[61] Jackson (n 58). But see Chapter 4 on meaningful human control, and Chapter 5 on reconceptualising 'critical functions' to the entire Joint Targeting process.

[62] See Chris Jenks, 'The Distraction of Full Autonomy and the Need to Refocus the CCW LAWS Discussion on Critical Functions' in Robin Geiß (ed.), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016) (advocating the same narrow focus discussed here, specifically to maintain a humanitarian focus).

[63] ICRC (n 51), 62-64. Again, see Chapter 5, which challenges this received wisdom on such a narrow focus.

[64] Scharre and Horowitz (n 30).

Three final points should be noted. First, there is a fourth possible dimension of autonomy: 'complexity of operational environment', which was presented in 2.2.3.2 as a determinant of reliable autonomy. This is not included in Scharre and Horowitz, yet it is implicit in some existing definitions,[65] and even singled it out as a separate 'dimension' of *weapons* autonomy.[66] Clearly, the operating environment will affect the level of risk to civilians and other protected persons and objects, all else being equal; thus, it should be part of the definition and, even more so, it should form one of the bases for deployment restrictions and pre-deployment precautions.

Second, while the above has conveniently divided each of the dimensions into three or so sub-categories, the reality is more complex and each one is actually a *spectrum* of infinite possibilities.[67] At the very least, there are other models that provide more than three levels, like Sheridan and Verplank,[68] OSD[69] and, more recently, Sharkey.[70] This will complicate the task of drawing objective boundaries between 'human-controlled and 'autonomous' systems, by offering more options that may conceivably fall into both categories, depending on how they are applied in practice.

Finally, the three (or four) dimensions are largely independent of each other and it is possible, for instance, to have a human totally 'out-of-the-loop', yet with such low 'machine complexity' as to effectively negate the concept of 'autonomy':[71] a landmine is a prime example. Furthermore, a given weapon system will simultaneously sit

---

[65] For example, Patrick Lin, George Bekey and Keith Abney, 'Autonomous Military Robotics: Risks, Ethics, and Design', *US Department of Navy, Office of Naval Research* (2008), 103 <http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil_fac> accessed 10 May 2018 (referring to the "capacity to operate in the *real-world environment*") (emphasis added). See also Wagner's definition, (text accompanying) n 49.

[66] See Christian Alwardt and Martin Krüger, 'Autonomy of Weapon Systems, *IFSH/IFAR Food for Thought Paper* (February 2016) <https://ifsh.de/file-IFAR/pdf_english/IFAR_FFT_1_final.pdf> accessed 10 May 2018.

[67] Scharre and Horowitz (n 30).

[68] TB. Sheridan and WL. Verplank, *Human and Computer Control of Undersea Teleoperators* (Man-Machines Systems Laboratory, MIT, 1978), 8.17-8.19 (explaining the 'Levels of Autonomy' model, which consists of ten levels of automation, going from complete human control through to complete machine control).

[69] Office of the Secretary of Defense, *Unmanned Aircraft Systems Roadmap 2005-2030* (DoD, 2005), D-10 (illustrating ten Autonomous Capability Levels, at Figure D-5).

[70] Noel Sharkey, 'Staying in the Loop: Human Supervisory Control of Weapons' in Nehal Bhuta et al. (eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (CUP, 2016), 28 (applying Sheridan and Verplank's model in a more explicit military context, and focusing on human-machine collaboration, specifically for designing systems with greater humanitarian impact. This leads the author to suggest five discrete levels of human supervisory control of weapons).

[71] Scharre and Horowitz (n 30), 7; Lin, Bekey and Abney (n 65), 103, authors' n 8.

somewhere along all three (four) dimensions; thus, its designation will depend on its *combined* positioning along all spectrums.

2.3.3 Towards a Working Definition of 'Weapons Autonomy' and LAWS

Taken together, the above points to 'weapons autonomy' as a feature that encompasses humans being *out of the loop*; in the (primary and/or secondary) *critical functions* of a weapon system; with those functions being performed by software controllers that are able to exercise *discretion*; in a potentially *complex and unstructured environment*. With these in mind, the following working definition of LAWS is proposed:

> A weapon system which, once activated, can select and engage targets without further human intervention and *usually* without any human pre-selecting those specific targets; *and*, in the process, to exercise discretion and self-direction to operate in a potentially complex and unstructured environment.[72]

At first sight, this may seem like a highly restrictive definition which, if all elements are to be *cumulatively* met, might overly narrow the scope of weapons autonomy. As mentioned above, however, an underlying concern is to maintain a humanitarian focus; together with the italicised words in the definition, this arguably provides appropriate flexibility. Thus, even in a targeted strike,[73] where the specific target *is* pre-selected by humans, residual machine 'discretion' on *choice of munition* and *timing of weapons release* will retain the autonomous nature of the deployment, thereby requiring appropriate precautions by deploying commanders.[74]

2.3.4 Three Clarifying Comments

2.3.4.1 The Technical Nature of LAWS 'Discretion'

First, 'discretion' is not used here to refer to anything like human-level intelligence, deliberative reasoning or true free will, these all presupposing certain human traits that machines inherently lack.[75] The software-based controllers of a LAWS will essentially be *deterministic* tools running on special-purpose computers that, however

---

[72] Maziar Homayounnejad, 'Assessing the Sense and Scope of 'Autonomy' in Emerging Military Weapon Systems', *TLI Think! Paper 76/2017* (2017), 15 <https://ssrn.com/abstract=3027540> accessed 10 May 2018.

[73] This is not a legal term, but is used in this thesis to denote an attack on preselected targets that have been developed through a human-led targeting process. See 5.2.5.1

[74] See also 2.4.3 on systems with lower levels of autonomy, which contradistinguish LAWS.

[75] See Chapter 3, especially 3.2.2 on the absence of human qualities in machines.

sophisticated, remain founded on the 'stored program' concept.[76] Namely, LAWS will be 'calculating machines', where "instructions entered by a human programmer are stored in the machine's memory and drawn upon to govern its operation".[77] Thus, 'discretion' is used here in a technical sense, in that the weapon system's controllers will a) collect (input) data, b) process it, and c) in accordance with that data and pre-programmed instructions, select one or more (output) options from a range of possible outcomes[78] *that may or may not have been foreseeable to deploying commanders*. The option chosen will always be a logical consequence of *ex ante* programming and sensed data, and this remains true even when machine learning comes into play.[79] Yet, as the opacity and learning ability of algorithms rise, and as the complexity of the battlefield increases – hence, the range of sensed data become less predictable – there may be the *appearance* of machine discretion. Certainly, autonomous systems will undertake stochastic (probability-based) reasoning,[80] which introduces greater uncertainty over their precise actions in a dynamic battlefield.[81] This is significant for those deploying LAWS, in that they may need to take *stronger, additional and earlier precautions* before any deployment, to safeguard against the risk of unintended engagements.[82]

2.3.4.2 Prioritising the Dimensions for Administrability

Second, there is the question of how administrable the working definition is, and which dimension of autonomy is the most useful. A potential problem identified by several authors is that it can be very difficult and subjective to draw a line between 'automated' and 'autonomous' systems,[83] especially based on any notion of apparent 'discretion'.

---

[76] William Aspray, 'Back to Basics: The Stored Program Concept' (1990) 27 IEEE Spectrum 51.

[77] McFarland (n 1), 15.

[78] In this regard, see Rebecca Crootof, 'The Killer Robots are Here: Legal and Policy Implications' (2015) 36 Cardozo Law Review 1837, 1854 (differentiating 'autonomous' from 'automated' systems by including in her definition of LAWS that the weapon system's critical actions will be "based on conclusions derived from gathered information and preprogrammed constraints…").

[79] McFarland (n 1), 16 (noting that while it is not immediately obvious, even a learning machine essentially executes instructions formulated by its developer. There is of course "an extra layer of abstraction between the developer and the weapon firing", which consists of new rules and algorithmic changes that originate not in the developer's mind, but in the data on which the system is subsequently trained. This extra layer of abstraction complicates the process of matching specific (attack) outcomes to specific (developer) commands, but it does not alter the fact that both the algorithmic changes and the final act of weapons release will involve the machine logically "executing instructions formulated by its developer").

[80] Cummings (n 28).

[81] ICRC (n 51), 13.

[82] See Chapter 7, especially 7.3.5 and 7.3.6.

[83] For example, Crootof (n 78); Scharre and Horowitz (n 30).

They contend that while 'machine complexity' provides a useful framework for thinking, it ultimately does not lend itself to legal definition, and thus it should not be applied beyond simple thought experiments or working definition status.[84] More recently, States at the April 2018 Group of Governmental Experts (GGE) meeting have argued for a move away from the automated/autonomous distinction due to its limited practical utility, and for a stronger focus on the type and degree of human control over weapon systems.[85] The main focus of discussions, therefore, seems to be the absence of direct human control over the *narrow* critical functions,[86] though machine complexity can and should arguably remain a feature of academic analysis.[87]

2.3.4.3 The Role of Humans

The final clarifying comment concerns the role of the human. To be sure, humans will not be totally 'out-of-the-loop' in the use of any LAWS that may be fielded in the near-term. Rather, commanders will retain full control of a number of important variables, which will be programmed into the systems.[88] At a bare minimum, these will include target parameters,[89] geographical boundaries and temporal boundaries.[90] This leaves the weapon system to select and engage *specific* targets that fall within these general constraints;[91] save in the case of targeted strikes.[92]

---

[84] Ibid.; Email from Paul Scharre to Maziar Homayounnejad (16 August 2016), on file with author (noting that machine complexity is a useful discussion tool, especially when addressing the 'black box' problem, but difficult to draw the line between 'automated' and 'autonomous' in practice; and preferring a legal definition of LAWS that "hinges upon what the machine is doing, agnostic to how smart it is", for clarity and legal certainty).

[85] See the various State contributions at the Second GGE on LAWS, 9-13 April 2018 <https://www.unog.ch/__80256ee600585943.nsf/(httpPages)/7c335e71dfcb29d1c1258243003e8724?OpenDocument&ExpandSection=-2#_Section2> accessed 10 May 2018.

[86] 'Narrow' in this context referring to target *recognition* by the weapon system's sensory hardware and control software, rather than the broader targeting *process* discussed in Chapter 5.

[87] United Nations Institute for Disarmament Research (UNIDIR), 'The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches', *UNIDIR Resources, No. 6* (2017), 21-22 <http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf> accessed 10 May 2018 (discussing the 'sequencing' proposal of prioritising a human-centric approach, before specifying the critical tasks and, finally, having a tech-centric discussion).

[88] See Chapter 4 on meaningful human control.

[89] That is, the exact categories (types) of targets that may be engaged, such as 'tank' or 'attack helicopter'.

[90] Article 36, *Killing by Machine: Key Issues for Understanding Meaningful Human Control* (6 April 2015) <http://www.article36.org/wp-content/uploads/2013/06/KILLING_BY_MACHINE_6.4.15.pdf> accessed 10 May 2018.

[91] Scharre and Horowitz (n 30). This includes discretion over the secondary critical functions.

[92] Where, as mentioned above, human commanders will pre-select the specific target, and machine discretion is limited to the secondary critical functions, to achieve the mission objective while minimising civilian harm.

Only within these limiting parameters will a LAWS be capable of offering genuine military utility, by enabling commanders to remain accountable for its actions and responsible for the overall outcome of operations in their own area of command and control. Concretely, such limiting parameters enable commanders to: fulfil mission objectives, in compliance with LOAC norms and mission Rules of Engagement, with full situational awareness,[93] and in pursuit of the broader strategic/political and military purpose of the operation.[94]

Accordingly, humans will remain in the 'wider loop' of (strategic and operational) control, while the weapon system operates with *relative* autonomy within the (tactical) 'narrow loop'.[95] As a corollary, "many key targeting decisions will…be made in *earlier phases* of the targeting cycle and at *locations further removed* from the intended strike site".[96] Thus, deliberative human reasoning on the actions of a LAWS will potentially be made in a more abstract setting, with less accurate knowledge of concrete threats or specific civilian risks. Again, this may require that commanders take *stronger, additional and earlier precautions* in their deployment and use of LAWS at the operational level, all else being equal.

2.4 Weapon Systems Likely to Emerge as LAWS

Horowitz points out that there are generally two kinds of weapon systems: munitions and platforms.[97] **Munitions** are physical and non-returnable/inherently one-way weapons designed to destroy a single target;[98] examples include missiles, bombs and rifle rounds. By contrast, **platforms** are inherently returnable systems that launch other munitions;[99] examples include combat aircraft, tanks and warships. Both types of weapon system can be autonomised with the integration of appropriate sensors,

[93] This includes the status and movements of own and allied forces, enemy forces, and the civilian population.

[94] I am grateful to Wolfgang Richter for pointing out these linkages.

[95] AIV and CAVV 'Autonomous Weapon Systems: The Need for Meaningful Human Control' *Report No. 97 AIV/No. 26 CAVV* (AIV/CAVV, October 2015). See Chapter 5. For a concrete application, see 7.2.3.4.

[96] Jeffrey S. Thurnher, 'Means and Methods of the Future: Autonomous Systems' in Paul AL. Ducheine, Michael N. Schmitt and Frans PB. Osinga (eds.), *Targeting: The Challenges of Modern Warfare* (TMC Asser Press, 2016), 178 (emphasis added).

[97] Horowitz (n 29), 94-97.

[98] Ibid.

[99] Ibid. The author points to a potential third category (operational system), but this is unlikely to be developed in the near-term, both for technical feasibility and military command and control reasons.

processing hardware and control software; either by original design, or via retrofit. Accordingly, the kind of weapon system that can be expected to emerge as 'autonomous' in the near-term is the 'wide-area search-and-attack' loitering *munition* and (drone) *platform*.[100] These will operate either as standalone units,[101] or as part of a swarm,[102] where collective behaviours can bring yet more dramatic and disruptive change to military operations.[103] Presently, a rather rudimentary autonomous munition exists in the Israeli *Harpy*. This detects and engages specific radar-emitting objects, with the option of negative visual confirmation,[104] and all within tight spatial and temporal boundaries within which deploying commanders believe lawful targets exist.[105] That said, the *Harpy* consists of many of the same technologies in today's semi-autonomous homing munitions, but with greater range and aerial persistence;[106] thus, it is mostly 'LAWS by usage'.[107]

2.4.1 Development of Standalone LAWS

Scharre notes that much of the technology for basic autonomous loitering munitions has existed for decades, but has not been widely developed or fielded.[108] A notable exception was the *Tomahawk Anti-Ship Missile* (*TASM*), which was in service in the US Navy from 1982 to 1994.[109] However, the *TASM* was retired and never actually fired in battle because Navy commanders were reluctant to launch a high-cost, limited-supply and *non-returnable* munition without concrete evidence that an actual warship was in the loitering area.[110] The advent of *recoverable* drone platforms arguably

---

[100] Scharre and Horowitz (n 30); Horowitz (n 29).

[101] Ibid.

[102] David Hambling, *Swarm Troopers: How Small Drones Will Conquer the World* (Archangel Ink, 2015).

[103] Paul Scharre, *Robotics on the Battlefield Part II: The Coming Swarm* (CNAS, 2014) <https://s3.amazonaws.com/files.cnas.org/documents/CNAS_TheComingSwarm_Scharre.pdf> accessed 10 May 2018 (discussing how swarms of robotic systems can bring greater mass, coordination, intelligence and speed to the battlefield, thereby increasing the chance of gaining a decisive advantage over adversaries).

[104] Namely, there is the option to visually zoom-in on the radar-emitting object, compare the image with a database of known 'friendly' sites; if no match is recognised, the *Harpy* proceeds to dive-bomb into its target.

[105] Paul Scharre, 'Autonomy, "Killer Robots," and Human Control in the Use of Force – Part I', *Just Security* (9 July 2014) <https://www.justsecurity.org/12708/autonomy-killer-robots-human-control-force-part/> accessed 10 May 2018.

[106] Ibid.

[107] Horowitz (n 29) points out, at 92-94, that the operator's *usage* can make semi-autonomous weapons fully autonomous.

[108] Scharre (n 13), 49.

[109] Ibid., 53.

[110] Ibid., 54.

changes this: these can be sent on missions with less accurate intelligence, and if no specific targets are detected, the drone can return to base with no loss of capacity.[111] This alone can be expected to increase their appeal relative to autonomous munitions. That said, the latter is still likely to be fielded for targeted strikes on fixed objects, where the munitions can carry out a similar function to cruise missiles, but with machine discretion on the timing of attack.

Accordingly, we can expect future LAWS – both munitions and platforms – to build on the current state of the art: to be more sophisticated with stronger artificial intelligence (AI) and automatic target recognition capabilities;[112] and to have longer loitering times and greater loitering areas. This will enable the systems to engage a wider range of targets with improved target accuracy; for example, by loitering longer for more multisensory and cross-cueing opportunities.[113] Stronger AI will also endow systems with the intelligence to detect, recognise and mitigate some civilian risk; for example, by varying the exact timing of attack in accordance with circumstances on the ground.

In addition to these, platform-based LAWS can be expected to have even longer loitering capabilities and a *choice of munitions* in attack. This may include a variety of different blast radiuses, which the control software may be able to match to its immediate environment before weapons release (to further mitigate civilian risk);[114] and also less-lethal/non-lethal munitions, to warn civilians to flee and/or incapacitate combatants rather than kill them, should this be consistent with mission goals.[115]

2.4.2 Development of Swarms

Perhaps even before standalone systems, *swarming* munitions will be developed, fielded and deployed first, if only for cost and strategic reasons.[116] This seems all the more likely, given recent occurrences of makeshift swarm attacks in Syria and Yemen,

---

[111] Ibid., 56. As opposed to a munition, which must either engage a target, self-destruct or dump itself at sea.

[112] See 2.5.1-2.5.4 on AI and automatic target recognition.

[113] See 2.5.4.1 on multisensor approaches and cross-cueing.

[114] See 7.3.2.2 on minimising collateral damage.

[115] See 7.3.2.4 on providing effective advance warning.

[116] Hambling (n 102); Scharre (n 103), 16-18 (discussing the contrasting effects of Augustine's Law, where linear budget rises are met with exponential cost increases *versus* Lanchester's Square Law, where reducing the size and increasing the quantity of combat assets will enable US forces to 'double up' on attacking enemy units).

and the gathering State and non-State competition to develop swarm systems for future deployment.[117] There are even suggestions that swarming may become the mainstay of future wars.[118] Generally, swarms can be utilised in three ways: *offensive attack*; *defence*; or in a *support* role, like providing intelligence, surveillance and reconnaissance (ISR).[119] In all cases, they involve "large numbers of dispersed individuals or small groups coordinating together and [operating] as a coherent whole".[120] Within these, the *individual* agents follow simple rules, from which the swarm *collectively* exhibits emergent intelligence, and complex and unified behaviours.[121] Accordingly, attack swarms will offer two distinct advantages over unitary systems: first, there is a **quantitative** element in which large numbers of micro-drones aim to *saturate* and *overwhelm* the enemy; second, there is a **qualitative (collaborative)** element, which aims to *outsmart* the enemy with rapid and unpredictable manoeuvres.[122] This can be seen as a modern application of the centuries-old military doctrine of mass and manoeuvre.[123] However, even before it approaches the target, an attack swarm will be able to jam enemy radar, and it has the advantage of potentially avoiding radar altogether (because of the small size and limited airspeed of each micro-drone unit);[124] it may disaggregate to avoid detection and concentrated attack, before reaggregating at the last moment to take the enemy by surprise.[125]

---

[117] See Maziar Homayounnejad, 'Drone Swarming and the Explosive Remnants of War', *Opinio Juris* (19 March 2018) <http://opiniojuris.org/2018/03/19/drone-swarming-and-the-explosive-remnants-of-war/> accessed 10 May 2018.

[118] Aaron Mehta, 'DoD Weapons Designer: Swarming Teams of Drones Will Dominate Future Wars', *Defense News* (30 March 2017) <https://www.defensenews.com/smr/unmanned-unleashed/2017/03/30/dod-weapons-designer-swarming-teams-of-drones-will-dominate-future-wars/> accessed 10 May 2018 (describing the Loyal Wingman concept, and the move towards expendable weapons).

[119] Scharre (n 103).

[120] Ibid., 26.

[121] Ibid.

[122] Maziar Homayounnejad, 'Autonomous Weapon Systems, Drone Swarming and the Explosive Remnants of War', *TLI Think! Paper 1/2018* (2018), 8 <https://ssrn.com/abstract=3099768> accessed 10 May 2018.

[123] John Arquilla and David Ronfeld, *Swarming and the Future of Conflict* (RAND, 2000).

[124] Interview with Air Force Chief Scientist Gregory Zacharias, in Kris Osborne, 'Air Force Seeks Swarms of Attack Mini-Drones', *Scout Warrior* (10 May 2016) <https://www.wearethemighty.com/articles/air-force-seeks-swarms-of-versatile-mini-drones> accessed 10 May 2018.

[125] Ibid.

Human controllers will supervise missions and instruct the broader goals, while individual autonomous units manoeuvre and perform various tasks unaided at the micro level.[126] Thus, swarms may tend closer towards human-*on*-the-loop systems, where unlawful behaviours can usually be manually overridden. Conversely, where they are deployed in a communications-denied environment, swarms will need the capability to operate in fully autonomous mode for at least part of the mission. This is likely to be the case in symmetric (high-intensity) conflict, and is the principal motivation behind DARPA's[127] *Collaborative Operations in Denied Environment* (*CODE*) program.[128]

An example of swarming – albeit in an ISR context – is the recent testing of 103 *Perdix* micro-drones, which demonstrated collective decision-making, adaptive formation flying and self-healing, all with a human operator defining broad tasks but not instructing any of these specific behaviours.[129] An obvious benefit of ISR swarming is to fuse together images from multiple viewpoints, for more accurate intelligence. Crucially, this may carry over into lethal autonomy: one of the stated benefits of the *CODE* program is "[p]roviding multi-modal sensors and diverse observation angles *to improve target identification*".[130] From this and other program goals, Scharre concludes that *CODE* may act as a gateway into lethal autonomy.[131]

The US Navy's LOCUST program (Low-Cost UAV Swarming Technology) goes a step further than the *Perdix* test flights, and aims to utilise swarming munitions (*Coyote* drones[132]) for ship defences;[133] while those designed for offensive attack will have the

---

[126] Scharre (n 103), 26.

[127] The Defense Advanced Research Projects Agency <https://www.darpa.mil/> accessed 10 May 2018.

[128] See Jean Charles Ledé, 'Collaborative Operations in Denied Environment (CODE)', *DARPA Program Information*, <https://www.darpa.mil/program/collaborative-operations-in-denied-environment> accessed 10 May 2018.

[129] US DoD 'Department of Defense Announces Successful Micro-Drone Demonstration', *Press Release No. NR-008-17* (9 January 2017) <https://www.defense.gov/News/News-Releases/News-Release-View/Article/1044811/department-of-defense-announces-successful-micro-drone-demonstration/> accessed 10 May 2018.

[130] DARPA, 'Broad Agency Announcement: Collaborative Operations in Denied Environment (CODE)', *Strategic Technology Office: DARPA BAA-14-33* (25 April 2014), 6 (emphasis added) <https://www.fbo.gov/utils/view?id=260d113392090305f63b281cf20bfea9> accessed 10 May 2018.

[131] Scharre (n 13), 72-76.

[132] 'Coyote UAS', *Raytheon* <http://www.raytheon.com/capabilities/products/coyote/> accessed 10 May 2018.

[133] David Hambling, 'US Navy Plans to Fly First Drone Swarm this Summer', *DefenseTech* (4 January 2016) <https://www.defensetech.org/2016/01/04/u-s-navy-plans-to-fly-first-drone-swarm-this-summer/> accessed 10 May 2018.

advantage of saturating and overwhelming enemy defences, such that even counter-measures and heavy defensive fire will not prevent a few 'leakers' from getting through to take out their target.[134]

Swarming concepts also extend to larger platforms. In 2014, the US Navy successfully tested 13-boat swarms via a retrofit system, both for maritime defence and for offensive attack on hostile vessels.[135] More recently, a Chinese company attracted media attention by demonstrating a 56-boat swarm in the South China Sea, which could "provide an asymmetric advantage in a conflict with the United States".[136]

2.4.3 Distant-Future and Existing Autonomy

Of course, autonomy is not a static concept, but is an incrementally advancing capability. Thus, LAWS will continue to develop into much more advanced (bipedal) systems in the longer term, as current projects in both the public sector[137] and private sector[138] suggest. However, because of the more distant time horizons, these will be excluded from the scope of this thesis. Equally, there is a plethora of currently-fielded weapon systems that incorporate autonomy to a greater or lesser degree in various functions. Again, these will be excluded from the scope of this thesis, but for the reason that they already operate under existing LOAC. These systems include:

- **Remotely-piloted systems**, such as the *Predator* and *Reaper* drones, which have a 'man-*in*-the-loop' both for selecting and engaging targets.
- **Semi-autonomous weapons**, such as precision-guided munitions, which automatically engage specific targets that have been pre-selected by humans.

---

[134] Scharre (n 103); Homayounnejad (n 117); Homayounnejad (n 122).

[135] David Smalley, 'The Future is Now: Navy's Autonomous Swarm Boats can Overwhelm Adversaries', *Office of Naval Research News & Media Center* (5 October 2014) https://www.onr.navy.mil/en/Media-Center/Press-Releases/2014/autonomous-swarm-boat-unmanned-caracas> accessed 10 May 2018 (reporting on the Control Architecture for Robotic Agent Command and Sensing (CARACaS) system).

[136] Robert Beckhusen, 'Chinese Robo-Boats Swarm the South China Sea', *War is Boring* (31 May 2018) <https://warisboring.com/dozens-of-chinese-robot-boats-swarm-the-sea/> accessed 3 June 2018.

[137] For example, in 2013 the Defense Advanced Project Research Agency (DARPA) allocated $7 million to the *Avatar Project*, which aims to enable human soldiers to partner with a semi-autonomous bipedal robot surrogate. See DARPA, 'Justification Book Vol. 1: Research, Development, Test & Evaluation, Defense-Wide', *Department of Defense FY 2013 President's Budget Submission* <https://www.darpa.mil/attachments/(2G4)%20Global%20Nav%20-%20About%20Us%20-%20Budget%20-%20Budget%20Entries%20-%20FY2013%20(Approved).pdf> accessed 10 May 2018.

[138] For example, the *Atlas* prototype by Boston Dynamics now has "balance and whole-body skills to achieve two-handed mobile manipulation", to enable the robot to "manipulate objects in its environment and to travel on rough terrain". See 'Atlas: The World's Most Dynamic Humanoid', *Boston Dynamics* <http://www.bostondynamics.com/robot_Atlas.html> accessed 10 May 2018.

- **Automated weapon systems**, such as the (ship-borne) *Phalanx CIWS* and the (land-based) *Patriot* air and missile defence systems. These select and engage incoming threats, but usually on simple rules-based criteria that work in a predictable manner.[139]

2.4.4 A Three-Part Chronology

The foregoing approximates to Farrant and Ford's 'three-wave' taxonomy of LAWS development.[140] According to the authors, 'first-wave' autonomous weapons are 'point defence systems',[141] such as the *Phalanx CIWS* and the *Patriot*. As just mentioned, however, this thesis will classify these as 'automated' systems, whose predictability and simplicity of algorithms do not necessitate additional regulation, or any restatement of laws.

'Second-wave' systems are "exemplified by the spate of unmanned combat aerial vehicles currently under development",[142] the clearest examples being the Dassault *nEUROn* and the BAE Systems *Taranis*. These systems – which Scharre imagines as "a *Predator* drone [with] as much autonomy as a Google [driverless] car"[143] – will operate on the wide-area loitering concept, and are of the kind that this thesis will address. As Farrant and Ford point out, however, second-wave systems are not limited to aerial vehicles, but will also include a range of ground-based[144] and maritime[145] autonomous systems, which will loiter or patrol over relatively narrower areas.[146] Examining the development trajectory of these LAWS, Jenks explains there will be a

---

[139] Often, these systems can also be set to operate in **supervised-autonomous** mode, where there is a 'man-*on*-the-loop', ready to override the machine's operation in the event of any unintended engagements.

[140] James Farrant and Christopher M. Ford, 'Autonomous Weapons and Weapons Reviews: The UK Second International Weapon Review Forum' (2017) 93 International Law Studies 389.

[141] Ibid., 396. That is, they select and engage incoming threats from a single location, and are designed to protect a particular object, like a warship or a forward-operating base.

[142] Ibid.

[143] Scharre (n 13).

[144] US Army, *Robotic and Autonomous Systems Strategy* (US Army Training & Doctrine Command, March 2017); Matthew Cox, 'Army Eyes Autonomous Convoys to Prevent Future Casualties', *Military.com* (1 May 2018) <https://www.military.com/dodbuzz/2018/05/01/army-eyes-autonomous-convoys-prevent-future-casualties.html> accessed 10 May 2018 (explaining that the US Army's new modernisation strategy is to develop teams of manned and unmanned ground combat vehicles, specifically to give commanders the option of autonomous attack to reduce friendly casualties).

[145] UNIDIR, 'The Weaponization of Increasingly Autonomous Technologies in the Maritime Environment: Testing the Waters', *UNIDIR Resources, No. 4* (2015) <http://www.unidir.org/files/publications/pdfs/testing-the-waters-en-634.pdf> accessed 10 May 2018.

[146] Farrant and Ford (n 140), 397-98.

'crawl-walk-run' approach, initially focusing on anti-material targeting in uncluttered environments where the civilian risk factor is relatively low, or non-existent.[147] Accordingly, autonomous maritime systems are likely to be developed and fielded in the largest numbers at an early stage,[148] as evidenced by recent military spending figures,[149] development plans[150] and project milestones.[151] This will be followed by developments in aerial LAWS[152] and, lastly, ground-based systems.[153] Subsequently, LAWS development will expand into more complex environments to undertake a broader range of attack missions, and this includes ground-based systems.[154]

Finally, Farrant and Ford refer to 'third-wave' systems, which will incorporate significantly increased autonomy and machine learning capabilities, thus will only be viable in the longer-term. Citing the DoD's 2013 *Unmanned Systems Roadmap*,[155] the authors refer to a movement from autonomous mission *execution* to autonomous mission *performance*. The difference is that second-wave systems will be goal-directed within pre-programmed parameters, whereas third-wave systems will focus on broader mission outcomes. Importantly, mission outcomes can vary *during* a

---

[147] Chris Jenks, 'The Gathering Swarm: The Path to Increasingly Autonomous Weapons Systems' (2017) 57 Jurimetrics 341, 347-50.

[148] Ibid., 348-49 (noting the vast expanse of the ocean environment, which is inhospitable towards inhabited systems, unreliable for communication links and remotely-piloted systems, and is an inherently less risky environment for LAWS deployments, especially those going subsurface). See also UNIDIR (n 145), 2 and 4-6 on these drivers for maritime LAWS.

[149] Jenks, ibid., 348 (citing the US Navy's growing budget to research and procure maritime LAWS, from $146.2 million in 2015, to $232.9 million in 2016, to over $350 million in 2017; and noting that the latter figure is four times the amount allocated for unmanned ground vehicles).

[150] Patrick Tucker, 'The US Navy is Developing Mothership Drones for Coastal Defense', *Defense One* (1 June 2018) <https://www.defenseone.com/technology/2018/06/us-navy-developing-mothership-drones-coastal-defense/148671/?oref=d-dontmiss> accessed 13 June 2018 (referring to the US Navy's Strategic Roadmap for Unmanned Systems, which shows that "the Navy is pushing to develop and buy its drones faster, integrate them more aggressively in exercises…and work more closely with universities…in the design of new prototypes").

[151] See 'ACTUV "Sea Hunter" Prototype Transitions to Office of Naval Research for Further Development', *DARPA News and Events* (30 January 2018) <https://www.darpa.mil/news-events/2018-01-30a> accessed 10 May 2018.

[152] Jenks (n 147), 349.

[153] Ibid., 350 (arguing that this will be severely limited in the early stages to sentry robots in the simplest environments, such as demilitarised zones, where the risk of civilian presence is low and is outweighed by the advantages of autonomous attack in the case of large-scale invasion).

[154] See US Army (n 144), 2-11 (highlighting the US Army's near-, mid- and far-term priorities for robotic and autonomous systems). See also Cox (n 144) ("The Army hopes to have its first Robotic Combat Vehicle technology demonstrator ready by 2021 so it can help inform future designs of autonomous combat vehicles").

[155] US DoD, *Unmanned Systems Integrated Roadmap*: FY 2013-2038 (DoD, 2014).

mission, thereby requiring *autonomous changes* to human-set parameters.[156] Again, the longer time horizons will exclude these systems from the scope of the thesis.

To conclude, the autonomous systems that are likely to enter the battlespace and raise LOAC compliance issues in the near-term are 'second-wave' loitering munitions and platforms, operating either as standalone units or within swarms. While many of these will be maritime and aerial vehicles designed for uncluttered environments – the so-called *Predator* with Google car capabilities – this will not invariably be the case as States will also field a number of ground systems. These near-term LAWS will almost certainly target large military objects, but they may also be designed to engage enemy combatants. Initially, second-wave systems will likely be a step-up from cruise missiles (in a targeted strike) or sensor-fused weapons (in tactical combat). Over time, they can be expected to grow in sophistication, and be fit for deployment in increasingly complex operational environments.

2.5 Artificial Intelligence and Automatic Target Recognition

Having delineated the kinds of weapon systems likely to emerge as near-term LAWS, it is worth examining in more detail some of the key enabling technologies of autonomous attack. This will allow a clearer understanding of potential LAWS capacities and limitations, especially in relation to the rules and principles of targeting law (Chapters 6 and 7).

2.5.1 Artificial Intelligence

Artificial intelligence (AI) is the very foundation of machine autonomy and it forms the basis for numerous weapons features and technologies. Essentially, AI is a "software endeavour",[157] which enables machines to replace *some* human judgment in warfare.[158] Accordingly, it has broadly been defined as "the capability of a computer system to perform tasks that normally require human intelligence",[159] which it does via knowledge representation and reasoning, to act as a rational agent.[160]

---

[156] Farrant and Ford (n 140), 398-99.

[157] Defense Science Board, US DoD (n 37), 22.

[158] For example, to exercise battlefield 'discretion' (see 2.3.4.1) in areas that largely require automatic processing (see 4.2.1).

[159] Cummings (n 28), 2 (providing the examples of visual perception, speech recognition and decision-making).

[160] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd ed., Pearson, 2016), 2-5 (explaining 'rational agent' as one that "acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome").

2.5.1.1 Narrow AI versus *General AI*

AI is broadly divided into narrow ('weak') and general ('strong') AI. **Narrow AI** has a limited range of cognitive abilities and specialises in one particular task, where it often equals or exceeds human performance *at that specific task only*; examples include driving a car,[161] playing Go,[162] or even competing in an aerial dogfight.[163] By contrast, **general AI** (or AGI) "can be applied to problems *in many different domains*, as human intelligence can".[164] So far, virtually all applications of AI – both civilian and military – are narrow, and this will likely remain true for the foreseeable future.[165] Indeed, as the working definition of LAWS emphasises,[166] near-term systems will autonomise a narrow range of *combat* functions.[167] Namely, the *primary* critical functions of finding, tracking, prioritising, selecting and engaging targets; along with the *secondary* functions regarding the choice of munition and the timing of weapons release. As will be seen in 2.5.3, this narrowness of AI and its limitation to *precise* and *tangible* tasks in a narrow range of environments[168] is the main cause of brittleness. This in turn will demand human control and stringent precautionary measures.

2.5.1.2 Top-Down (Rules-Based) versus *Bottom-Up (Learning) Approaches*

Another vital distinction concerns 'top-down' (rules-based) *versus* 'bottom-up' (learning) approaches to AI. A **Top-down** system has the programmer defining not only the *problems* to be solved by the software, but also the *way* they are to be solved. This results in long computer codes, often based on simple *ex ante* rules, which tell the system what it *shall do* or what it *can conclude* in specific situations.[169] Not surprisingly, this approach is suitable for precise, well-defined tasks in a relatively simple environment, where all or most variables can be easily expressed in code.

---

[161] Paul Gao, Russell Hensley and Andreas Zielke, 'A Road Map to the Future of the Auto Industry', *McKinsey Quarterly* (October 2014).

[162] Dan Silver et al., 'Mastering the Game of Go Without Human Knowledge' (2017) 550 Nature 354.

[163] Nick Ernest and Kelly Cohen, 'Genetic Fuzzy Based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions' (2015) 6 Journal of Defense Management 139.

[164] Keith Frankish and William M. Ramsey (eds.), *The Cambridge Handbook of Artificial Intelligence* (CUP, 2014), (emphasis added).

[165] Katja Grace et al., 'When Will AI Exceed Human Performance? Evidence from AI Experts' (v3, 20 May 2018) <https://arxiv.org/pdf/1705.08807.pdf> accessed 10 June 2018.

[166] See 2.3.3.

[167] In addition to some non-combat tasks, like take-off and landing, navigation, and fault-detection.

[168] See 2.2.3.2.

[169] Alison Cawsey, *The Essence of Artificial Intelligence* (Prentice Hall, 1998), 29.

Automated weapons like air and missile defence systems[170] work in this way, because a) they are often deployed on a warship sailing the high seas, or on remote land near a forward-operating base, and b) they serve the sole purpose of defending said military objects against unambiguous and relatively clearly-defined incoming threats.[171] An example of a top-down code for a *Phalanx CIWS* may be:

IF (object sited <= 2000 metre radius) + (approaching at >= 200 mph) + (descending altitude)
THEN (aim) + (fire)
ELSE (remain on standby)

Further code may specify slightly more detailed tasks, though in all instances these remain relatively simple, and are often based on predictable rules rooted in the laws of physics.[172] Significantly, top-down/rules-based programming is limited by the constrained ability of humans to model complex environments,[173] and this clearly makes it unsuitable as the *primary* mode of programming a LAWS.[174]

This brings us to the **bottom-up** approach to AI, which relies primarily on **machine learning**; this is broadly defined as "the ability of a program to learn from experience, that is, to modify its execution on the basis of newly acquired information".[175] Accordingly, learning machines are *adaptive systems* and, as such, are potentially suitable for complex and dynamic battlefields. Domingos succinctly explains the difference between the two approaches:

---

[170] See (notes and text accompanying) nn 45 and 139 on the *Phalanx CIWS* and the *Patriot* missile battery.

[171] Namely, in an uncluttered environment such as the high seas, a speeding object moving and descending towards a warship may be deemed to only constitute an enemy threat (e.g. a cruise missile or an aircraft on the attack), thereby justifying an armed response in and of itself.

[172] For example, the *Phalanx* senses the speed and trajectory of the incoming cruise missile or aircraft, and it takes into account both windage and ballistic elevation to adapt the direction of its Gatling gun in a timely fashion. This keeps its stream of ammunition aiming towards the incoming threat. Likewise, a *Patriot* missile will sense the speed and trajectory of the incoming threat, and it will automatically alter its flight path to keep in line with the target.

[173] Boulanin and Verbruggen (n 19), 14.

[174] Though rules-based programming can still be used for routine functions, like autonomous navigation; or for any objective tasks, like attacking a GPS coordinate.

[175] See *Nature Reviews* Glossary <http://www.nature.com/nrg/journal/v10/n6/glossary/nrg2579.html> accessed 1 June 2016.

> Every algorithm has an input and an output: the data goes into the computer, the algorithm does what it will with it, and out comes the result. Machine learning turns this around: in goes the data and the desired result and out comes the algorithm that turns one into the other. Learning algorithms…are algorithms that make other algorithms. With machine learning, computers write their own programs, so we don't have to.[176]

A common form of machine learning is the **artificial neural network** (ANN), which is loosely modelled on the neuronal structure of the human brain; in particular, the cerebral cortex.[177] In a human context, this consists of billions of neurons (nodes) firing electrical charges at one another, each one generated by thoughts and perceptions. The result is a structural change in the brain, and new or updated 'rules' upon which it will perceive and classify future events.[178]

2.5.1.3 Machine Learning Methods

Similarly, an ANN learns by being fed data (e.g. images, or real-world visuals of tanks) from which it derives conclusions (e.g. recognising a 'tank'). With a critical mass of data, the ANN develops detailed algorithmic rules, which enable it to draw correct conclusions in the field.[179] This can be done through **supervised learning**, whereby human agents feed the machine labelled input data *and* the desired result,[180] after which learning is effectuated from both correct predictions and backpropagation of errors.[181] Namely, when an incorrect output is provided (e.g. misidentifying a bus for a tank) an error value is generated, which reflects the divergence between the correct

---

[176] Pedros Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (Allen Lane, 2015), 6.

[177] Bharat Bhosal, 'Curvelet Interaction with Artificial Neural Networks' in Subana Shanmuganathan and Sandhya Samarasinghe (eds.), *Artificial Neural Network Modelling* (Springer International, 2016), 117.

[178] Ibid. (noting that ANNs mimic this biological process with an artificial structure consisting of *processing elements* (nodes); *interconnections* between those elements/nodes; an *activation function* (rule), which transforms input into output (inside a node); and a *learning function*, which manages the weights of input-output pairs. Accordingly, ANNs 'learn from example' or, more correctly, from input-output data samples. This entails a strengthening or weakening of connections between the many nodes in the system via statistical correlation, until the desired behaviour is realised. Knowledge is therefore encoded in the strength of the neural connections).

[179] Ibid.

[180] The desired result is some variable X (in this case, a tank) that is designated as the target for prediction, explanation or inference. Moreover, the values of X in the dataset (the tanks in the image samples) constitute the "ground truth" values for learning. See David Danks, 'Learning', in Frankish and Ramsey (n 164) 151, 154.

[181] Juergen Schmidhuber, 'Deep Learning in Neural Networks: An Overview' (8 October 2014) <https://arxiv.org/pdf/1404.7828.pdf> accessed 13 May 2018.

and incorrect answer.[182] The error value is then backpropagated through the ANN, which updates the input-output weights/connections between the nodes, to correctly output 'tank' when an image of a tank is next shown.[183] Thus, the algorithm learns to *generalise* its training data to new instances of 'tank'. Supervised learning is the most common method thus far, and is suitable for training an algorithm in the laboratory.

Another method is **unsupervised learning**, where unlabelled datasets with no known answers are fed into the system.[184] The system then searches for hidden patterns, and this often leads to 'clustering algorithms', where the inputs are separated into "natural groups" exhibiting some discernible characteristic.[185] For example, a sample of battle-damage assessment data may reveal that collateral damage tends to be lower with delayed-fusing.[186] The program would cluster the inputs into 'low' and 'high' collateral damage, without explicitly being introduced to this concept.

Finally, there is **reinforcement learning**, where the program interacts with the real-world, and is supplied with reward functions.[187] Every time the system's action leads to a desired result a positive reward is provided to reinforce that behaviour; whereas an undesired result leads to a negative reward or 'error-signal'.[188] The algorithm then decides which actions led to the specific reinforcements and adjusts itself accordingly, without any further human input.[189]

It is expected that most future ANNs will utilise a combination of all three learning methods, depending on context and circumstances,[190] and this is likely to be true with LAWS.[191]

---

[182] Yann LeCun, Yoshua Bengio and Geoffrey Hinton, 'Deep Learning' (2015) 521 Nature Review 436.

[183] Ibid.

[184] Russell and Norvig (n 160), 694.

[185] Ibid.; Danks (n 180), 154.

[186] This is where the bomb fuse is timed to detonate a certain period after impact.

[187] Russell and Norvig (n 160), 695.

[188] Ibid.

[189] Ibid. Note, however, this can be complicated by the credit-assignment problem, where the program finds it difficult to link a specific action with the relevant reward function.

[190] LeCun, Bengi and Hinton (n 182).

[191] For example, a LAWS will undergo supervised learning in relation to known target categories, where there is an abundance of training data. Unsupervised learning may uncover new intelligence from existing data, and support the development of new targets, new warfighting tactics, or new ways to mitigate civilian risk. Reinforcement learning may enable a LAWS to develop its own warfighting tactics and defensive/evasive manoeuvres against specific types of threats, similar to playing a game.

2.5.1.4 Modes of Machine Learning

In addition, algorithms can be trained 'online' or 'offline', using sequential or simultaneous (batch) data inputs, respectively.[192] **Online learning** sees the program learning incrementally from new external data, as and when these are collected by its sensors.[193] This enables the system to exploit growing volumes of data, where these are unknown, or represent changes to existing datasets.[194] In a LAWS context, this would mean learning from interactions in a live battlefield, which would enable the system to:

- adapt to new backgrounds and climatic conditions;
- refine its warfighting tactics, for greater targeting accuracy and precision;
- keep pace with changing enemy tactics in real-time, for example, by developing more effective evasive/defensive manoeuvres or new methods of attack; and
- generalise new target sets that may have been missed during the formal targeting process.

Offline learning is where the program is restricted to its original training data.[195] For a LAWS, this largely refers to laboratory training, which takes place during product development, and will be subject to human scrutiny via verification and validation procedures before the algorithmic changes are finally integrated into the system.

It can be expected that in a warfighting context, offline learning will be preferred for the greater predictability and human control that it affords. However, as online data enable algorithms to remain up-to-date and combat-effective, they are still likely to be recorded by a LAWS on deployment, before being reviewed by human programmers and selectively uploaded during system updates. It is also possible that limited forms of online learning will be allowed, for example, to refine the machine's targeting functions but without undertaking whole new tasks or methods of warfare that were not anticipated by deploying commanders.

---

[192] Claude Sammut and Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning* (Springer, 2010), 74, 736-43; Pat Langley, *Elements of Machine Learning* (Morgan Kaufmann, 1996), 9.

[193] Russell and Norvig (n 160), 752-53.

[194] Ibid.

[195] Sammut and Webb (n 192); Langley (n 192).

2.5.2 Deep, Convolutional and Recurrent Neural Networks

Building on the above (ANN) approach to machine learning is a more recent development, **deep learning**, which is effectuated through a **deep neural network** (DNN).[196] This has become increasingly viable with a) the exponential rise in computing power, and b) the availability of ever-larger datasets, partly as a result of the proliferation of internet-enabled devices.[197] The outcome of this twin phenomena has seen newer ANNs superseding previous ones by a factor of ten or more.[198] With so many nodes and connections between them, AI researchers are now able to take another cue from the biological brain: namely, to organise those nodes into distinct, hidden hierarchical layers, with each successive layer identifying higher level abstractions.[199] Figure 2.4, below, illustrates this model. After feeding input data (images of a face), the first hidden layer of the network scans individual pixels, their brightness and colours, to identify edges and shadows. The second hidden layer identifies combinations of edges, to recognise specific features like 'eyes', 'nose' or 'mouth'. This continues until the output layer, which classifies the object ('face').[200]

In a military context, the same approach may be used to train a network to identify military objects, like tanks. Thus, after feeding input data (images of a tank), the first hidden layer may scan individual pixels, to identify edges and shadows. The second may identify combinations of edges and shadows, to recognise specific features like 'tank treads' or 'gun barrel'. This will continue until the output layer, which should classify the object as a 'tank'.

---

[196] LeCun, Bengi and Hinton (n 182).

[197] Domingos (n 176).

[198] Jeremy Hsu, 'Biggest Neural Network Ever Pushes AI Deep Learning', *IEEE Spectrum* (8 July 2015) <http://spectrum.ieee.org/tech-talk/computing/software/biggest-neural-network-ever-pushes-ai-deep-learning> accessed 13 May 2018.

[199] 'Artificial Intelligence: Rise of the Machines', *The Economist Briefing* (9 May 2015).

[200] Bart Haar Romeny, 'Tutorial: Deep Learning in Human and Computer Vision' (30 August 2017) <http://bmia.bmt.tue.nl/people/BRomeny/Courses/Taipei2017/index.html> accessed 13 May 2018.

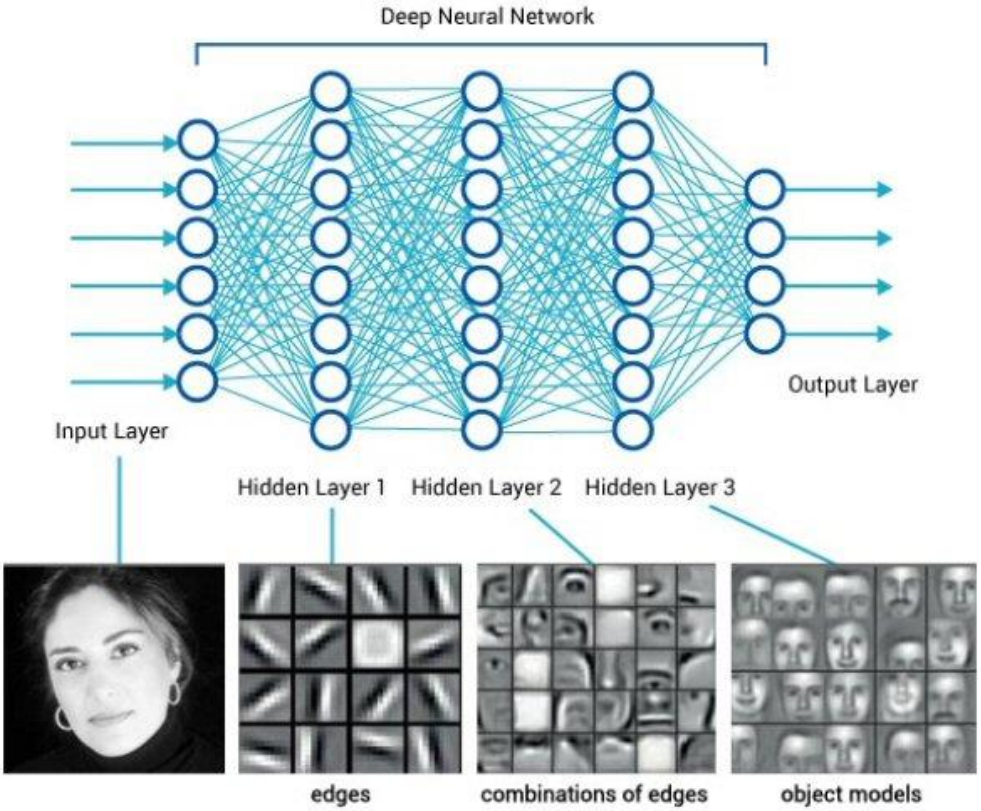

Figure 2.4: Deep neural network and its hidden layers. Source: Romeny (n 200).

It is the use of these interlinked layers that gives this approach to machine learning its 'deep' characteristic. Adding more layers between input data and output makes the network more complex, enabling it to handle more sophisticated tasks with greater detail and accuracy.[201] The result is significantly lower error rates in such tasks as object, voice and (potentially) activity recognition by DNNs, relative to smaller ANNs.[202] Consequently, DNNs – or, in the case of image recognition, **convolutional neural networks** (CNNs)[203] – are now able to detect subtle distinctions that humans typically miss.

[201] Scharre (n 13), 87.

[202] Kaiming He et al., 'Deep Residual Learning for Image Recognition' (10 December 2015), 2 <https://arxiv.org/pdf/1512.03385.pdf> accessed 13 May 2018 (detailing an image recognition algorithm with 152 layers, which achieved an error rate of 3.57% compared with the human error rate of 5.1%).

[203] CNNs are essentially DNNs that are optimised for image recognition.

More recently, facial appearances have been correlated with human emotional states, resulting in a greater prediction accuracy achieved by a machine learning algorithm, than by human subjects.[204] This may have implications on the battlefield, for example, in detecting persons *hors de combat*.[205]

All of this, however, relates to individual faces and objects in isolation. Karpathy and Fei-Fei describe the next stage of object recognition as a computer vision system that is able to recognise *multiple* objects in a single picture, and to label specific parts of that picture using a **recurrent neural network** (RNN).[206] Their ultimate goal is to generate dense, free-form image descriptions that simultaneously recognise the contents of an image *and the overall context*;[207] hence, (partially) addressing the problem of AI brittleness (see 2.5.3).

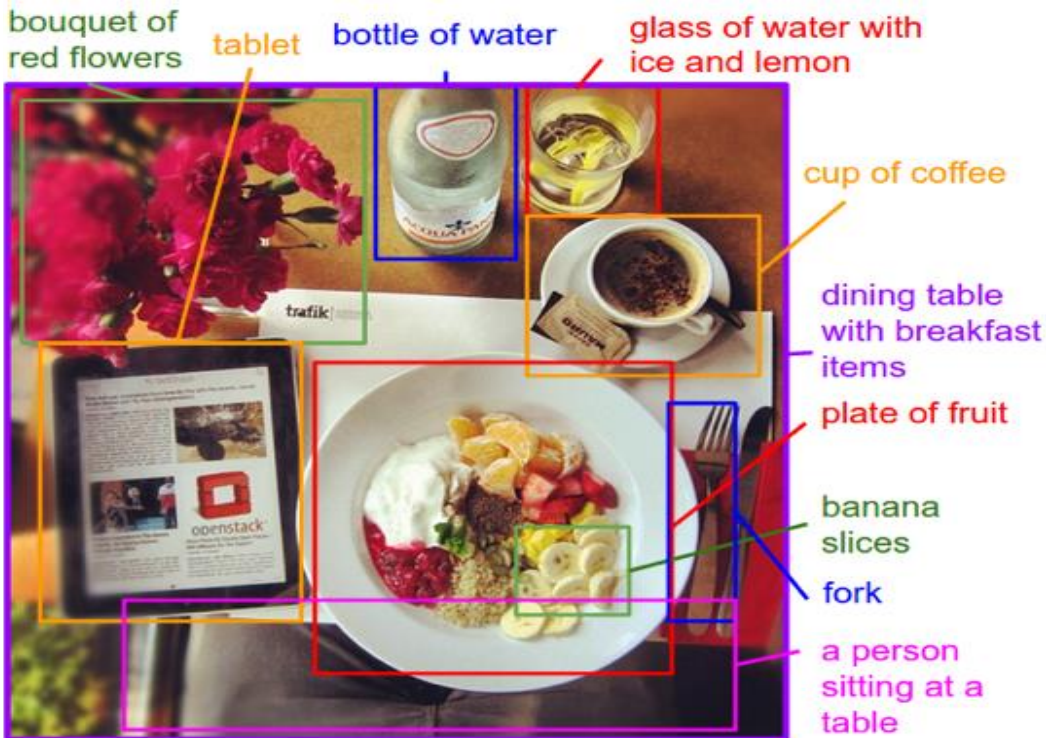

Figure 2.5: A RNN's take on breakfast. Source: Figure 1, Karpathy and Fei-Fei (n 206).

[204] Carlos F. Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M. Martinez, 'Facial Color is an Efficient Mechanism to Visually Transmit Emotion' (2018) 115 Proceedings of the National Academy of Sciences 3581 <http://www.pnas.org/content/pnas/115/14/3581.full.pdf> accessed 21 May 2018.

[205] See 6.5.2.4.

[206] Andrej Karpathy and Li Fei-Fei, 'Deep Visual-Semantic Alignments for Generating Image Description' (CVPR, 2015) <https://cs.stanford.edu/people/karpathy/cvpr2015.pdf> accessed 13 May 2018.

[207] Ibid.

Figure 2.5, above, illustrates the authors' concept. The system is shown a single image, in response to which it identifies individual items such as the fork, a person sitting at the table, a plate of fruit and even banana slices *within* that plate. It also identifies the 'bigger picture' (dining table with breakfast items), although the authors found this more abstract capability for pinpointing the overall context was imperfect and prone to error.[208]

Significantly, deep learning is not limited to images but is a *general-purpose* technique for statistical pattern-recognition. In principle, this can apply to any activity for which there are sufficient datasets,[209] and it has been argued that the "potential military applications are virtually unlimited".[210] In particular, DNNs in their various guises may enable such capabilities as accurately identifying enemy combatants, high-value individuals from a variety of angles, or even persons *hors de combat*; precise and accurate weapons release; activity recognition and prediction of hostile actions;[211] and navigating highly complex terrain.[212] These capabilities are all critical in situations where milliseconds count,[213] and where a mistake in target identification, the detection of protected status, or collateral damage estimation may have dire humanitarian consequences. Accordingly, DNNs will form an integral part of future automatic target recognition (see 2.5.5.4), as well as intelligence and target development (see 5.5.3).

However, machine learning is not without its problems,[214] and there are challenges such as 'overfitting' to the data,[215] as well as a dearth of useful data.[216] Moreover, the emphasis on replicating the human brain structure does not necessarily mean that DNNs 'think' like humans. On the contrary, the networks display brittleness and have

---

[208] Ibid.

[209] For some examples of its diverse applications, see Yann LeCun, Yoshua Bengio, Y and Geoffrey Hinton, 'Deep Learning' (2015) 521 Nature Review 436, 436.

[210] Farrant and Ford (n 140), 399.

[211] Chandler P. Atwood, 'Activity-Based Intelligence: Revolutionizing Military Intelligence Analysis' (2015) 77 Joint Force Quarterly 24.

[212] Farrant and Ford (n 140), 399.

[213] Both for getting inside the enemy's OODA loop, and for cancelling/suspending attacks upon seeing civilians enter the effects area.

[214] See International Panel on the Regulation of Autonomous Weapons (iPRAW), 'Focus on Computational Methods in the Context of LAWS', *"Focus on" Report No. 2* (November 2017), 11-12.

[215] Ibid. This is where simple mathematical representations fit too closely to training data, and are not generalizable or practically useful.

[216] Ibid.

some strongly counterintuitive properties that remind us of the essentially *technical* nature of any LAWS that will utilise them.

2.5.3 The Inherent Brittleness of AI

As noted in 2.5.1.1, current and near-future AI are narrow, in that they have a limited range of cognitive abilities and specialise in one or a few tasks. This in turn is a significant cause of **brittleness**, whereby systems do *precisely* but *only* what they are programmed/trained to do.[217] Thus, while they typically exceed human performance in narrow domains – e.g. object recognition, or fast, accurate and precise responses to clearly-defined stimuli – they lack the 'common sense' to understand the broader context, and to adjust their outputs and behaviours accordingly. Consequently, when unexpected events occur, or the environment or context for action changes, errors and system failures often follow.[218] Unlike humans, autonomous systems cannot discard their 'instruction book' and use 'common sense' to adapt to the situation at hand. Indeed, any programming based on such a premise will necessarily lack the precision and tangibility to be machine-executable.[219] Accordingly, as a primer for the legal analysis in Chapters 3, 6 and 7, the following will highlight some of the main sources of AI brittleness.

2.5.3.1 Brittleness in Context

In a LAWS context, brittleness can manifest itself in several mutually non-exclusive ways, including the following.

- Perceiving an object or a signal that normally triggers an attack response, but without recognising qualitative factors, or taking into account other contradictory signals. The result is often a false positive target identification, thus a risk of unlawful engagement. This is a drawback of many current target recognition technologies.[220]
- Being thrown off-track by minor perturbations, like a change in the weather[221] or irrelevant changes in the image or physical appearance of an object.[222]

[217] Scharre (n 13), 145.
[218] Ibid., 146.
[219] See 2.2.3.2.
[220] See 2.5.5.2.
[221] See 2.5.5.1.
[222] See 2.5.6.

- Failing to take correct action in a particular instance for lack of quality training data.[223]
- 'Perverse instantiation' of final goals, whereby the system fulfils the narrow criteria of its programmed objectives, but in unanticipated ways that violate the programmer's broader intentions,[224] and possibly the LOAC rules.[225] This is because algorithms solve via constrained optimisation, whereby one or a few objectives (those that have been programmed) are focused upon at the exclusion of all else.[226]

Brittleness is expected to be a pervasive problem affecting LAWS and their associated AI. While this generally does not mean the systems should be banned, it does counsel in favour of human control and suitable precautionary measures.

2.5.3.2 Brittleness and the Risk of Learning 'Wrong Lessons'

It was noted in 2.5.1.4 that machine learning can occur offline (in the laboratory) or online (in the battlefield), and that the former will be preferred, for the greater control that it affords. Namely, to avoid situations where a LAWS may learn the 'wrong lessons', which in turn may lead to target generalisation towards civilian objects, or even a perverse instantiation of final goals. The civilian world has numerous examples of 'wrong lessons' being learnt. These include the following.

- The Microsoft *Tay* chatbot, which learnt the worst of human nature after 24 hours on Twitter.[227]
- The *Playfun* algorithm, which was given the goal of 'not losing' at Tetris, and technically achieved this by pausing the game indefinitely just before the last (losing) block fell in place.[228]

---

[223] See 2.5.5.3.

[224] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (OUP, 2014), 146-49 (providing some hypothetical examples in the context of an AGI). See also (notes and text accompanying) nn 228-231, for some real-world examples.

[225] See (notes and text accompanying) nn 232 and 244-245, for some IHL/LOAC-relevant examples.

[226] Stuart Russell, 'Of Myths and Moonshine', *Edge* (14 November 2014) <https://www.edge.org/conversation/the-myth-of-ai#26015> accessed 13 May 2018 (noting that where the system is optimising a function of *n* variables, but the objective depends on just a subset of these (the constrained variables), it will often set the remaining (unconstrained) variables to extreme values. Yet, if any of those unconstrained variables is important for IHL/LOAC compliance, the solution may be perverse and unlawful).

[227] Daniel Thomas, 'Microsoft Pulls Twitter Bot Tay after Racist Tweets', *Financial Times* (24 March 2016) <https://www.ft.com/content/8ba60bc4-f1c0-11e5-aff5-19b4e253664a> accessed 13 May 2018.

[228] Tom Murphy VII, 'The First Level of Super Mario Bros. is Easy with Lexicographic Orderings and Time Travel…After That it Gets a Little Tricky' (1 April 2013)

- An evolution strategy algorithm tasked with maximising its score at the Q*bert video game, and achieving this by learning to cheat.[229]
- An early AI tasked with developing 'rules of thumb' for gameplay. One of these (H59) learnt to maximise its score 'fraudulently', by finding other high-scoring rules and putting itself down as the originator, thereby taking credit without adding any value.[230]
- Fears have now emerged that market trading algorithms may learn to collude and spoof each other on-the-fly, to manipulate market outcomes and maximise profits.[231]

Clearly, the last example is the most concerning, as it poses a risk of significant economic harm. Yet, it is likely that online learning in a military context would be even more perilous, due to the safety-critical context; the canonical example being a LAWS that might 'learn' it is quicker and more efficient to defeat the enemy by killing all humans in sight, rather than just those meeting target parameters.[232] While this may be a stylised example, the real-world instances of 'wrong lessons' all counsel against the idea of online learning by a LAWS in principle, perhaps subject to narrow functional and/or contextual exceptions;[233] even then, with meaningful human control.[234]

2.5.3.3 Transparency, Opacity and the 'Black Box' of Machine Learning

It is worth distinguishing some *known* from *unknown* elements in machine learning. Assuming offline learning will be the predominant form of training a LAWS algorithm, the first **known** element will be *data input*, in that the programmer will

<https://www.cs.cmu.edu/~tom7/mario/mario.pdf> accessed 13 May 2018 (noting that *Playfun* remained totally unremarkable at actual gameplay in Tetris).

[229] Patryk Chrabaszcz, Ilya Loshchilov and Frank Hutter, 'Back to Basics: Benchmarking Canonical Evolution Strategies for Playing Atari' (24 February 2018) <https://arxiv.org/pdf/1802.08842.pdf> accessed 13 May 2018.

[230] Douglas B. Lenat, 'EURISKO: A Program That Learns New Heuristics and Domain Concepts' (1983) 21 Artificial Intelligence 61.

[231] See Ian Allison, 'When Intelligent Algorithms Start Spoofing Each Other, Regulation Becomes a Science', *International Business Times* (29 June 2016) <http://www.ibtimes.co.uk/machine-learning-markets-when-intelligent-algorithms-start-spoofing-each-other-regulation-becomes-1567986> accessed 13 May 2018.

[232] Michael Schmitt, 'Regulating Autonomous Weapons Might be Smarter than Banning Them', *Just Security* (10 August 2015) <https://www.justsecurity.org/25333/regulating-autonomous-weapons-smarter-banning/> accessed 13 May 2018.

[233] Potentially acceptable are the first two bullet points (and the defensive point in the third), in 2.5.1.4.

[234] See Chapter 4.

deliberately select and feed-in the data, training scenarios and reward functions from which he wants the program to learn. These choices will invariably be constrained by the LOAC rules and they eliminate the need for algorithms to determine their own data sources or target parameters; save in the case of online learning, which in the near-term will likely be circumscribed.

The second known element is the *probability thresholds* set by the programmer. As the machine learning features noted above suggest,[235] data pass through a stochastic system, which transforms inputs into *probable* outputs. Concretely, the program finds the probability distribution that explains the training data, then it runs inference on that distribution to label new inputs. Whether the probability/confidence attached to a new input is high enough to proceed with a certain categorisation (e.g. positive identification of a 'tank') or action ('engage') is predetermined by the programmer.[236]

Yet, despite these human-controlled elements, the overall decision-making process in machine learning is fundamentally a 'black box',[237] and this represents its greatest **unknown** element. As noted above, machines write their own algorithms to achieve a human-defined result,[238] but it is often unclear *which* characteristics within the data sets are chosen to be the focus of learning, and how their respective *weightings* are determined.[239] This may lead to unanticipated outputs, which no human in-the-loop would have authorised. For example, in the case of supervised learning of 'tank' images, a program that weights the tank treads disproportionately may risk identifying agricultural or construction vehicles with caterpillar tracks as a 'tank'. Without a human in-the-loop, the system may then open fire on these civilian vehicles.

Moreover, *utility functions* are often used to optimise certain criteria;[240] for example, to maximise target hits within a ten-metre radius, or to minimise collateral damage.

---

[235] See n 178.

[236] Igor Kononenko and Matjaž Kukar, *Machine Learning and Data Mining: Introduction to Principles and Algorithms* (Horwood, 2007).

[237] W. Nicholson Price, 'Black-Box Medicine' (2015) 28 Harvard Journal of Law & Technology 419, 432-34; Gary Coglianese and David Lehr, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era' (2017) 105 Georgetown Law Journal 1147, 1159.

[238] Note and text accompanying (n 176).

[239] Coglianese and Lehr (n 237), 1159 ("we cannot really know what precise characteristics any machine-learning algorithm is keying in on").

[240] Ibid.; UNIDIR (n 87), 11.

This clearly aims to replicate human discretion, particularly in situations where value judgments are made in order to resolve conflicting interests; the application of IHL/LOAC in armed conflict being a canonical example.[241] However, while utility functions are written by the programmer and nominally within his control, they are very difficult to specify in the abstract, due to the need for *precision* and *tangibility*[242] in the face of *bounded rationality*.[243] Namely, the programmer selects the objectives and utility criteria, but cannot cover every eventuality or predict exactly *how* these will be applied in the field. The result could be any number of unforeseen consequences, which may amount to a perverse instantiation of final goals, due to the inherent brittleness and lack of 'common sense' in AI systems.[244] For example, a LAWS tasked with maximising target hits, while minimising civilian casualties, might do so in a way that collaterally destroys a piece of infrastructure (e.g. a water treatment plant). This may directly kill fewer civilians than any of the alternative attack options, but with potentially greater reverberating effects (longer-term civilian deaths, due to the loss of vital infrastructure). Conversely, a human in-the-loop may have been better-placed to consider the full effects, as circumstances unfolded, and in the particular context of that mission. Accordingly, Russell and Norvig argue:

> [W]ith AI systems…we need to be very careful what we ask for, whereas humans would have no trouble realizing that the proposed utility function cannot be taken literally.[245]

To summarise, programmers control data input and provide a utility function to the best of their abilities, and within constraints. Beyond that, the machine learning process is a 'black box', which remains *inscrutable* (difficult to grasp and validate its processes) and often *non-intuitive* (difficult to understand its output). Moreover, such is the complexity of the decision-making process that even a forensic analysis often fails to deconstruct and specify its elements.[246]

---

[241] See 3.2.2.2.1 on the need for metacognition, and 7.2.2 on the problematic compliance with proportionality.

[242] See 2.2.3.2.

[243] Herbert A. Simon, *Models of Man: Social and Rational* (Wiley, 1957).

[244] See (notes and text accompanying) nn 224-226 on perverse instantiation.

[245] Russell and Norvig (n 160), 1037.

[246] Price (n 237).

To be sure, programmers may try to build-in software tools for *observability* of internal processes and *directability* if these make errors,[247] so that systems are more transparent and predictable, and end-users have greater control over their behaviours. A recent example is 'Explainable AI' (XAI), which is an ongoing DARPA project. This aims to produce a suite of tools and techniques that enable commanders and other users to understand, appropriately trust, and manage an emerging generation of AI machine partners.[248] If its project goals are met, XAI may offer a solution to the black box problem, though it is too early to draw any conclusions.[249] For now, the twin problem of inscrutability and non-intuitiveness continues to aggravate the long-standing issue of AI brittleness; an issue to which we will return in 2.5.6-2.5.7 below,

2.5.4 Automatic Target Recognition

Automatic target recognition (ATR) refers to the "automatic (unaided) processing of sensor data to locate and classify targets".[250] In principle, this is essential for enabling a LAWS to distinguish legitimate military objectives from civilians and other protected persons and objects, and from background clutter. In this sense, ATR will be essential for the 'Find, Fix, Track' parts of the kill chain, as well as 'Assess', so a LAWS may determine whether it needs to re-engage.[251]

2.5.4.1 Standard ATR Approaches

ATR packages comprise potentially vast combinations of hardware and software, which are designed for particular operational environments.[252] Examples include video cameras, light (LIDAR), electro-optical/infrared (including thermal imaging) sensors; acoustic (sonar and ultrasonic) sensors; radar and electromagnetic sensors; DNNs, and the Global Positioning System (GPS). These constellations enable **object recognition**

---

[247] UNIDIR, 'Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies', *UNIDIR Resources, No. 5* (2017), 6 <http://www.unidir.org/files/publications/pdfs/safety-unintentional-risk-and-accidents-en-668.pdf> accessed 13 May 2018.

[248] Specifically, the goal of XAI is to provide an audit trail, documenting which factors and features weighed into an algorithm's decision to produce a given output. See David Gunning, 'Explainable Artificial Intelligence (XAI)', *DARPA Program Information* (10 August 2016) <http://www.darpa.mil/program/explainable-artificial-intelligence> accessed 13 May 2018.

[249] For a project update, see David Gunning, 'Explainable Artificial Intelligence (XAI)', *DARPA Program Update* (November 2017) <https://www.darpa.mil/attachments/XAIProgramUpdate.pdf> accessed 13 May 2018.

[250] Bruce J. Schachter, *Automatic Target Recognition* (3rd ed., SPIE Press, 2018), 1.

[251] See 5.3.2.1.

[252] Schachter (n 250).

of large military objects by 'nature' or by 'location',[253] via a quantitative assessments of easily-recognisable characteristics, like image, size, shape, sound, heat, velocity, material content and GPS coordinates.[254] These are derived from a 'multisensory phenomenology' of a number of predefined target signatures,[255] which are cross-cued with data on geolocation and distance from target.[256] Once these characteristics are reconciled with a template stored in the system's target identification library, the weapon system is able to lawfully attack.[257] In the case of multiple target detections, the ATR will also prioritise between targets based on strict predefined criteria.[258] Thus, consistent with 2.2.3.1 on the sense-think-act paradigm, current ATR relies on pattern recognition techniques and it lacks any deliberative qualities; hence, some commentators argue that it enables not autonomous, but *automated* targeting at most.[259]

Consequently, ATR systems are mostly employed to detect large and well-defined military objects, such as tanks, aircraft, warships, submarines, or any radar-emitting objects. Robot sentry weapons like the Samsung *SGR-A1* can detect human targets via heat signatures, but cannot currently determine whether these are military personnel or civilians.[260] Accordingly, robot sentries have only been deployed in very simple environments, such as demilitarised zones where civilians should generally not be

---

[253] See 6.5.3 for a more detailed application of these legal criteria.

[254] Wagner (n 50), 1391-92; Michael Lewis, Katia Sycara and Paul Scerri, 'Scaling Up Wide-Area-Search-Munition Teams' (2009) 24 IEEE Intelligent Systems 10.

[255] Schachter (n 250), Chapter 5 (explaining that there are three multisensory phenomenologies:

1) 'Multisensor', where more than one of the same sensor modes (e.g. two infrared devices) detect the same target.
2) 'Multilook', where the same sensor detects the same target from different aspects/angles, often to build a 3D image.
3) 'Multimode fusion', where sensors of different modalities detect the target (e.g. the fusing of acoustic and electro-optical signals), similar to how humans use the five senses to perceive the world).

[256] Wendell H. Chun and Nikolaos Papanikolopoulos, 'Robot Surveillance and Security' in Bruno Siciliano and Oussama Khatib (eds.), *Springer Handbook of Robotics* (Springer-Verlag, 2016), 1613.

[257] The exact characteristics and number of signatures that must be reconciled for 'positive identification', as well as the prescribed minimum 'confidence threshold', will vary from one target to the next, and will likely be determined by prior testing and evaluation. See Alan Backstrom and Ian Henderson, 'New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews' (2012) 94 International Review of the Red Cross 483, 495, 508-12.

[258] Boulanin and Verbruggen (n 19), 24.

[259] Ibid.

[260] Moreover, the capacity of sentry robots to detect surrender is very basic (both arms held high), and will almost certainly be inadequate for complex and chaotic environments: Boulanin and Verbruggen (n 19), 25.

present.[261] There are suggestions that ATR may in the near-term be able to recognise traditional combatants,[262] and while this is feasible in principle, it will require significant improvements in ATR technologies.[263]

2.5.4.2 Cooperative and Non-Cooperative Targets

There is an important distinction between 'cooperative' and 'non-cooperative' targets.[264] Cooperative targets emit a signal, which makes them easier to detect; examples include any radar-emitting objects, or submarines emitting acoustic signals. These are typically detected with *passive* sensors such as radar receivers, which distinguish them from civilian and friendly forces via IFF (Identification Friend or Foe) systems, before homing-in to destroy the target.[265] By contrast, non-cooperative targets do not broadcast any signal; examples include vehicles with their radars switched off, submarines running silently; tanks, mobile missile launchers or artillery pieces. These have to be detected with *active* sensors, such as cameras, which pick up light and process these for image recognition; radar transmitters, which send out electromagnetic energy and 'see' reflected signals from the target; or sonar sensors, which send out sound waves and 'hear' echoes bouncing off the target.[266]

LAWS will generally operate more effectively and reliably against cooperative targets. However, this is not guaranteed in cluttered environments where there may be large numbers of radar-emitting objects in close proximity, which do not all transmit appropriate IFF squawks. In such circumstances, target indication (see below) and/or multisensory phenomenologies will be needed for detection-confirmation, similar to how the *Harpy* operates.[267]

2.5.4.3 Target Indication versus *Target Identification*

ATR can perform both *target indication* and *target identification*. The latter is the main capability described above, which would potentially enable a LAWS to operate

---

[261] Ibid., 24.

[262] See 6.5.2.

[263] Boulanin and Verbruggen (n 19), 24-25.

[264] Scharre (n 13), 84-86.

[265] The Israeli *Harpy*, explained in 2.4 above, is an example of a current weapon system that uses this approach.

[266] See David Blacknell and Hugh Griffiths (eds.), *Radar Automatic Target Recognition (ATR) and Non-Cooperative Target Recognition (NCTR)* (The Institute of Engineering and Technology, 2013).

[267] Recall from 2.4 that the *Harpy* detects radar-emitting objects with the option of negative visual confirmation.

autonomously. By contrast, target indication is a mode of operation of a radar that enables the user to discriminate between a *potential* target and background clutter.[268] This is further subdivided into stationary target indication (STI) and moving target indication (MTI).[269] On its own, target indication generally needs a human in-the-loop, to provide contextual reasoning before launching an attack.

Another meaning of 'target indication' is seen in the optical and acoustic sensing technologies, which generate initial leads for further action. A recent example is the *Boomerang III,* which pinpoints the location and direction of hostile gunfire in less than one second,[270] and which has also been miniaturised into a wearable device for individual soldiers.[271] Both detection systems have been successfully deployed by US and UK ground forces.[272]

Like STI and MTI, optical/acoustic sensing generally needs a human in-the-loop for contextual reasoning. However, the technology can arguably be integrated in a LAWS, to generate initial leads for 'tipping and cueing', in order to hasten target identification.[273]

2.5.4.4 GPS Guidance Systems

One specific component of ATR – its GPS guidance system – can, in some circumstances undertake the bulk of the target recognition task. GPS pinpoints the exact coordinates of a weapon system's current geographical location, and the location of external sites and landmarks.[274] Thus, it is vital for navigation (be that manned, remote or autonomous), and is a primary source of 'Position, Navigation and Timing'

---

[268] Melvin L. Belcher, Jr. and James A. Scheer, 'Radar System Implementation' in Jerry C. Whitaker (ed.), *The Electronics Handbook* (2nd ed., CRC, 2005) 1820.

[269] Ibid. (explaining that MTI indicates the presence of a moving object (like an aircraft) amongst stationary background objects, like hills; while STI focuses on the intrinsic characteristics of a target and its clutter, such as their relative sizes).

[270] 'Boomerang III: State-of-the-Art Shooter Detection', *Raytheon* <https://www.raytheon.com/capabilities/products/boomerang> accessed 18 May 2018.

[271] 'Boomerang Warrior-X: Wearable Shooter Detection System for Soldiers', *Raytheon* <https://www.raytheon.com/capabilities/products/boomerang_warriorx> accessed 18 May 2018.

[272] Gemma Carroll, 'On Target with Boomerang III: Acoustic Sensing Technology', *Army Technology* (30 May 2012) <https://www.army-technology.com/features/featuredssi-boomerang-iii-dstl/> accessed 18 May 2018.

[273] Tipping and cueing is where a predefined action (like gunfire) automatically directs other sensors towards the initial source, to progress from target indication to detection-confirmation and full target identification.

[274] See 'GPS: The Global Positioning System' <http://www.gps.gov/> accessed 18 May 2018.

(PNT)[275] data. Accordingly, GPS enables **location recognition** for engaging *fixed* targets such as a bridge, a permanent command post, or any other building that is prioritised for attack;[276] and for avoiding fixed protected objects, like medical facilities or cultural property.[277] As an additional check that the correct target is being selected, GPS/PNT data can be combined with the object recognition capability of the ATR.[278]

However, being transmitted from satellites that are 12,000 miles from Earth, GPS signals are notoriously weak and fragile, especially at low altitudes.[279] Hence, they are vulnerable to jamming, hacking and spoofing,[280] and this is what enabled Iran's mid-air commandeering of a US stealth surveillance drone in 2011.[281] Moreover, the return to Great Power Competition suggests that near-peer adversaries will continue to step up their investments in anti-access/area denial (A2/AD) networks.[282] This counsels against any over-reliance on satellite-based communications, especially in the context of symmetric (high-intensity) conflict, if the strategic value of LAWS is to be retained.[283]

2.5.4.5 Vision-Based Guidance Systems

That said, recent advances in (vision-based) electro-optical/infrared (EO/IR) guidance systems are sharpening location recognition capabilities, while eliminating the risk of GPS hacking.[284] These systems use a scene-matching correlator, which pinpoints the weapon system's location and its intended target by comparing real-time images of the current terrain (and fixed objects) with an onboard database of stored maps and

---

[275] Kevin M. Coggins, 'Position, Navigation and Timing and What it Means for the Soldier', *DoD Armed with Science* (27 February 2016) <http://science.dodlive.mil/2016/02/27/staying-on-course-positioning-navigation-and-timing-pnt-and-what-it-means-for-the-soldier/> accessed 18 May 2018.

[276] That is, a 'targeted strike' by location. See 6.5.3.1.

[277] See 6.5.3.4.1, 6.5.3.4.3 and 6.5.3.4.4.

[278] Similar to how the *Harpy* operates. See 2.4.

[279] 'Comments of the National PNT Advisory Board', *Jamming the Global Positioning System – A National Security Threat: Recent Events and Potential Cures* (4 November 2010), 3 <http://www.gps.gov/governance/advisory/recommendations/2010-11-jammingwhitepaper.pdf> accessed 18 May 2018.

[280] Ibid., 4-6.

[281] Jeff Hecht, 'Did Iran Capture US Drone by Hacking its GPS Signal?' *New Scientist* (16 December 2011); 'Iran Shows 'Hacked US Spy Drone' Video Footage', *BBC News* (7 February 2013) <http://www.bbc.co.uk/news/world-middle-east-21373353> accessed 18 May 2018.

[282] See 1.2.2.

[283] But see 7.2.3.4, specifically on new developments, which aim to augment or replace GPS signals in the face of A2/AD networks.

[284] See 'Smart Weapons: The Vision Thing', *The Economist* (3 December 2016).

images.[285] Accordingly, EO/IR scene-matching enables **secure fixed object recognition**[286] for targeted strikes in GPS-denied areas.

2.5.5 The Brittle Nature of ATR and Potential Solutions

The above suggests that current ATR systems are proficient at detecting a *limited* number of military objects in a *limited* number of scenarios. Accordingly, the technology remains brittle and this is true for at least four reasons.

2.5.5.1 Sensitivity to Environmental Conditions

First, systems can be very sensitive to battlefield smoke and weather conditions: as soon as these degrade the machine's sensory perception, false-alarm rates have been found to increase significantly.[287] Thus, any LAWS equipped with near-term ATR cannot be deployed in all operational environments, or for extended periods of time. That said, there are three possible solutions. First, reliability rates can be improved with distributed sensors via swarming,[288] and indeed this is one of the stated benefits of the DARPA *CODE* program.[289] Second, in some specific missions, like a targeted strike on a fixed object, EO/IR scene-matching may cut through the weather problems. Finally, the broader problem may change in future with developments such as 'ghost imaging'.[290]

2.5.5.2 Narrow Domains versus *the Broader Context*

Second, being a narrow form of AI, ATR systems merely *recognise* predefined target types based on pre-programmed criteria,[291] but generally cannot discern the surrounding context.[292] This is crucial, as a recognisable object like a rifle may have a

---

[285] Ibid.

[286] That is, to recognise specifically targetable buildings, infrastructure or terrain, but without relying on fragile and deniable communications links.

[287] James A. Ratches, 'Review of Current Aided/Automatic Target Acquisition Technology for Military Target Acquisition Tasks' (2011) 50 Optical Engineering 1.

[288] Prithviraj Dasgupta, 'Distributed Automatic Target Recognition Using Multi-Agent UAV Swarms', *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (8-12 May 2006), 479 (describing a distributed search with 'tipping and cueing' behaviour. Thus, when one micro-drone senses a potentially targetable object, others swarm towards it to collectively perform ATR and detection-confirmation, thereby benefiting from more sensors, data-collection and data-processing capabilities).

[289] See 2.4.2.

[290] Ghost imaging is an application of quantum physics, which derives an artificially-generated, but vastly improved holographic image of an object that might be two or so miles away on a smoky battlefield. See 'The Newest Thing in Quantum Imaging', *DoD Armed with Science* (3 January 2014) <http://science.dodlive.mil/2014/01/03/the-newest-thing-in-quantum-imaging> accessed 14 May 2018.

[291] Boulanin and Verbruggen (n 19).

[292] Scharre (n 13).

number of contrasting uses, from perfectly innocuous ones[293] through to offensive attack. Assessing which of these applies in a given scenario requires an understanding of context and intuition, which currently only humans have.[294] As Schachter points out, "determining the gist of a scene and doing so as well as an experienced image analyst is proving difficult to automate".[295] The situation is even more acute where foreign cultures and customs present additional sources of misunderstanding.[296]

Such narrowness of object recognition highlights the fact that ATRs perform *quantitative* interpretations of observed data, but they do not understand the underlying reality or context that produces those data. This counsels against any over-reliance on ATR and it points to the need for human input,[297] either to supervise the system with a man-on-the-loop, or to restrict its autonomous operation to relatively simple environments and tasks.[298]

2.5.5.3 Scarcity of Quality Data

The third reason for ATR brittleness is a chronic data-shortage problem. While the above limitations may all be resolved with a critical mass of training and test data, in many cases these are unique to each mission scenario.[299] For example, training samples need to include appropriate data not just on each target category, but also on mission-specific variations like different backgrounds and climatic conditions.[300] Because of the potentially transient nature of such variables, this problem is often difficult to resolve in practice. Online learning may seem to offer a potential solution, but in the near-term the results are arguably too unpredictable for a safety-critical mission.[301]

---

[293] For example, for recreational sport, lawful self-defence, or simply for cultural reasons, such as celebratory gunfire, which is not uncommon in parts of the Middle East.

[294] Cummings (n 28).

[295] Schachter (n 250), 245.

[296] Wagner (n 50), 1392-93 (discussing a hypothetical counter-insurgency operation, where a confluence of factors leads to ambiguity. As soldiers approach a house where they suspect an insurgent may be hiding, child inhabitants playing with a ball kick it towards the gate and run after it. Adult male inhabitants carrying a *kirpan* dagger for religious reasons scream at the children to stay away from the gate, and begin running towards them. Such a scenario may be easy to interpret for a human soldier, but an ATR may perceive the rapid approach of multiple males carrying a dagger to be a sign of impending attack).

[297] Ibid.; Schachter (n 250). See also Chapter 4.

[298] Ibid.; Scharre (n 13), 146.

[299] Boulanin and Verbruggen (n 19).

[300] Ibid. Consistent with Wagner's *kirpan* scenario, Schachter (n 250) adds, at 244, that learning local cultural traits is also vital if an ATR is to target persons (even if only in a reactive or self-defence capacity), to avoid misconstruing innocuous behaviours.

[301] Boulanin and Verbruggen (n 19).

Moreover, rare and unexpected combinations, as above, are – by definition – not amenable to machine learning.

2.5.5.4 Structural ATR Weaknesses and the Proposed TRACE Solution

A particular problem with trying to detect targets in a cluttered environment is the abundance of signals and the performance-degrading effects on ATRs of decoys and background traffic.[302] Even where genuine signals can be separated from the noise, this often requires significant computing power, which thus far has been unviable for mobile platforms.[303] An alternative is to use Synthetic Aperture Radar (SAR), which is an active sensor that sends out multiple radar pulses and collects the returning signals to construct an image of objects within range.[304] However, the images are grainy and difficult enough for humans to interpret.[305] Thus, for ATR systems designed for autonomous attack, current SAR techniques will offer little input value.[306]

In this connexion, DARPA is currently trying to solve the ATR problem under its Target Recognition and Adaptation in Contested Environments (TRACE) program. This aims to "develop algorithms and techniques that rapidly and accurately identify military targets using radar sensors on manned and unmanned tactical platforms".[307] Concretely, TRACE intends to combine the most advanced DNNs and other machine learning techniques with SAR, to enable clearer images of a target to be constructed from standoff ranges; for these to be amenable to ATR detection and classification amid background traffic and decoys; and, crucially, to achieve all of this with lower computing power and energy consumption than current systems.[308] DARPA summarises the TRACE program goals as follows.[309]

---

[302] John Gorman, 'Target Recognition and Adaptation in Contested Environments (TRACE)', *DARPA Program Information* <https://www.darpa.mil/program/trace> accessed 14 May 2018.

[303] Ibid.

[304] Schachter (n 250), 63.

[305] Ibid., 92.

[306] Scharre (n 13), 86; Schachter, ibid.

[307] DARPA, 'Broad Agency Announcement: Target Recognition and Adaptation in Contested Environments (TRACE)', *Strategic Technology Office: DARPA BAA-15-09* (1 December 2014), 6 <https://www.fbo.gov/utils/view?id=d68e7ccabd593f2f46b2328fa18dc6db> accessed 14 May 2018.

[308] John Keller, 'DARPA TRACE Program Using Advanced Algorithms, Embedded Computing for Radar Target Recognition', *Military and Aerospace Electronics* (24 July 2015) <http://www.militaryaerospace.com/articles/2015/07/hpec-radar-target-recognition.html> accessed 14 May 2018.

[309] DARPA (n 307), 6.

(1) Military target recognition on *low-power* airborne platforms.

(2) Low false-alarm rates for targets deployed in *complex* environments.

(3) *Rapid learning* of new targets with *sparse* or *limited* measured training data.

Whether or not this will actually be used for autonomous attack, it will certainly provide an important building block for LAWS development. For if the program goals are realised, TRACE will construct ATR systems that can identify non-cooperative targets like tanks, mobile missile launchers and artillery pieces, with at least human-level proficiency.[310] This will technically remove the need for humans in the critical functions narrow loop.

Scharre also notes that the TRACE program is likely to harness the recent advances in computer vision and image recognition, noted above.[311] If successful, this would enable yet more accurate identification of non-cooperative targets at closer range, potentially including uniformed combatants.

2.5.6 Fooling the ATR with Adversarial Examples

However, the accuracy and reliability of image recognition depends not only on the sophistication of computer vision systems, but also on their *robustness*. On the one hand, current DNNs and CNNs are highly advanced in relation to undistorted images, where the systems exhibit human-level[312] or super-human-level[313] performance in object classification. On the other hand, those same systems get rapidly confused when there are image degradations caused by, for example, additive noise, heat, or other environmental distortions.[314]

---

[310] Scharre (n 13), 88.

[311] Ibid. See 2.5.2.

[312] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, 'ImageNET Classification with Deep Convolutional Neural Networks, *Proceedings of the 25th International Conference on Neural Information Processing Systems* (2012) 1097.

[313] Kaiming He et al. (n 202); Kaiming He et al. 'Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNET Classification', *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)* (2015) 1026.

[314] Robert Geirhos, et al., 'Comparing Deep Neural Networks Against Humans: Object Recognition When the Signals Get Weaker' (21 June 2017) <https://arxiv.org/pdf/1706.06969.pdf> accessed 16 May 2018.

2.5.6.1 The Nature of Adversarial Examples

More specifically, systems are easily 'spoofed' by minor deliberate changes to the image,[315] which are largely imperceptible to humans, but enough to fool the system into misclassifying the object with a high degree of confidence (see Figure 2.6, below). In some instances, it takes no more than a single-pixel alteration to successfully spoof the network.[316] It is true that the ATR of a LAWS will not typically view static images, but will sense three-dimensional persons and objects in the physical world, where a multilook approach with viewpoint shifts will support detection-confirmation.[317] Yet, even this is no guarantee against spoofing: as Athalye et al. demonstrate, carefully-designed yet minor changes to physical objects can *consistently* fool a computer vision algorithm, despite those objects being viewed from multiple angles (see Figure 2.7).[318]

Figure 2.6: Adversarial images. The left and right columns look identical to humans, but were perceived differently by the DNN. The undistorted images (left column) were correctly identified by the network. Adversarial static (middle column, with 10x magnification) was added to create the distorted images (right column), causing the DNN to identify all three as an 'ostrich'.

Source: Figure 5(b), Szegedy et al. (n 315).

Figure 2.7: Adversarial objects. Actual 3D printed turtle classified as a rifle, and a baseball classified as an espresso, even when viewed from multiple angles.

Source: Adapted from Figures 13 & 14, Athalye et al. (n 318).

---

[315] Christian Szegedy et al, 'Intriguing Properties of Neural Networks' (19 February 2014) <https://arxiv.org/pdf/1312.6199.pdf> accessed 16 May 2018.

[316] Jiawei Su, Danilo Vasconcellos Vargas and Kouichi Sakurai, 'One Pixel Attack for Fooling Deep Neural Networks' (22 February 2018) <https://arxiv.org/pdf/1710.08864.pdf> accessed 16 May 2018.

[317] As opposed to two-dimensional images, which offer a single viewpoint. Note also the possibility to utilise distributed sensors via swarming, for enhanced ATR accuracy. See (note and text accompanying) n 288.

[318] Anish Athalye et al., 'Synthesizing Robust Adversarial Examples' (30 October 2017) <https://arxiv.org/pdf/1707.07397.pdf> accessed 16 May 2018.

These distorted images and objects are known as 'adversarial examples',[319] as they "exploit [DNNs'] vulnerabilities to trick them into confidently identifying false images".[320] Importantly, such distortions are relatively robust, as many of the vulnerabilities are common to most DNNs, regardless of their parameters, *or even their training data*.[321] This means that an attacker need not know the specific internal structure of a target DNN, nor have any other proprietary knowledge about it in order to fool the system. In a twist of irony, improved DNNs and CNNs that further enhance object classification will not necessarily resolve the spoofing problem, as it is often that same technology which finds more subtle ways to fool the improved algorithms into misclassifying images and objects.[322]

Not surprisingly, such spoofing tactics will be a significant problem in adversarial settings such as armed conflict, where opposing parties have an incentive to send each other confusing signals. Whether such actions amount to a lawful ruse of war, a perfidy, or a violation of any other obligation will be briefly addressed in 6.5.2.1.2. For now, it is worth noting that this vulnerability casts doubt on using the current class of visual object recognition and other similar DNNs for military applications, unless commanders can take adequate active precautions to mitigate the associated risks.

2.5.6.2 The Apparent Unavoidability of Adversarial Risk

The underlying problem is that complex software artefacts appear to classify images, as well as sound and other activity,[323] in highly counterintuitive and fragile ways. In the case of DNNs, these are strongly *linear* at the micro-level, even if non-linear at the macro-level, and this turns out to be a major source of vulnerability.[324] It becomes particularly significant when the networks operate on high-dimensional manifolds

---

[319] Not to be confused with the meaning of 'adversarial' in armed conflict.

[320] Scharre (n 13), 180.

[321] Szegedy et al (n 315); Su, Vargas and Sakurai (n 316); Anh Nguyen, Jason Yosinski and Jeff Clune, 'Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images', *Computer Vision and Pattern Recognition (CVPR '15)* (IEEE, 2015) <https://arxiv.org/pdf/1412.1897.pdf> accessed 16 May 2018.

[322] Miles Brundage et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (February 2018) <https://maliciousaireport.com/> accessed 16 May 2018.

[323] Corey Kereliuk, Bob L. Sturm and Jan Larsen, 'Deep Learning and Music Adversaries' (16 July 2015) <https://arxiv.org/pdf/1507.04761.pdf> accessed 16 May 2018 (explaining how a song-interpreting DNN was fed with adversarial audio signals, which sounded like nonsense to humans but was confidently interpreted as music by the DNN).

[324] Ian J. Goodfellow, Jonathan Shlens and Christian Szegedy, 'Explaining and Harnessing Adversarial Examples' (v3, 20 March 2015), 2-3 <https://arxiv.org/pdf/1412.6572.pdf> accessed 16 May 2018.

consisting of millions (or even billions) of dimensions, where there are always pockets that are misclassified and which can be easily exploited to fool the system.[325] Consequently, DNNs can falsely identify objects from the addition of meaningless static in ways that humans generally do not, if such static is carefully designed to exploit the network's vulnerabilities.

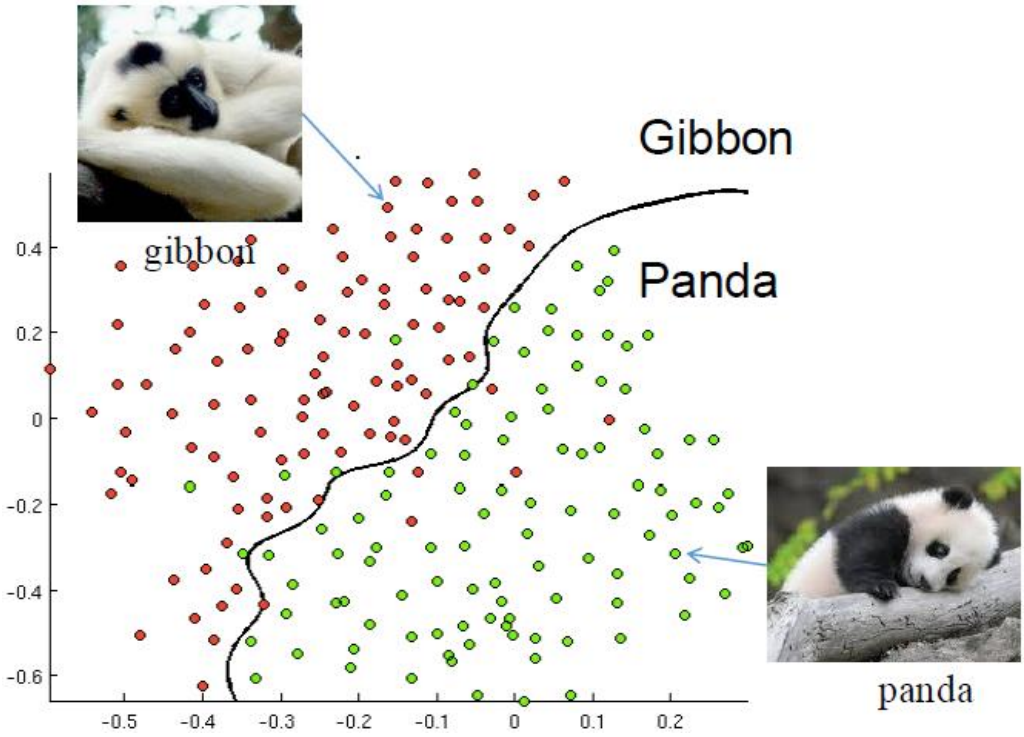

Figure 2.8: A simplified two-dimensional manifold. Source: Figure 14, JASON (n 325).

To illustrate via a simplified example,[326] Figure 2.8 depicts a two-dimensional manifold consisting of either 'panda' images (green data points) or 'gibbon' images (red data points). If the DNN is trained on this data alone and asked to predict whether a given data point is likely to be a panda or a gibbon, it will draw a line (or 'decision boundary') through the data points to derive two image domains. The network would then predict that new points to the north/west of the boundary are likely to be gibbons, while those to the south/east are likely to be pandas; recognising also that some overlap exists. Asked to predict the most likely image of a gibbon, the DNN would put it 'infinitely far to the north/west', even though the network has no data on any points

[325] JASON, *Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to the DoD* (JSR-16-Task-003, January 2017), 28-31 <https://fas.org/irp/agency/dod/jason/ai-dod.pdf> accessed 16 May 2018.

[326] The following is adapted from Scharre (n 13), 183-85.

that far away. Worse still, because of the extreme northerly position of the hypothetical data point, the DNN would classify it as a gibbon *with very high confidence*. Hence, the simple linear representation of data at this micro-level, as the network assumes the further one moves upwards/leftwards, the more likely the image is of a gibbon; adversarial examples exploit this vulnerability.

However, as Clune points out, real-world images – which DNNs classify more accurately than humans – are "…a very, very small, rare subset of all possible images".[327] The above represents weakness on the extremes, in an actual space of virtually infinite possible images. Accordingly, the real problem is more complex than Figure 2.8 depicts, and it exists in a very *high-dimensional* manifold where boundaries between numerous image domains are not always clear, or even correct.[328]

2.5.6.3 Can an AI Be Inoculated Against Spoofing?

Numerous researchers have successfully used 'adversarial training' to tweak a DNN, to recognise a specific false image.[329] However, this mostly "expands the instruction book", and it does not comprehensively train the DNN to "recognise the underlying pattern in the instruction book and to process new examples correctly".[330] Thus, Goodfellow, Shlens and Szegedy found that while DNNs typically became more robust with adversarial training, the remaining error rate and misclassification of false images was still too high to be acceptable for safety-critical systems.[331] Concretely, adversarial training builds robustness against the specific *kinds* of attacks used during training, but not towards whole new categories of attacks.[332] Again, the reason is that the space of all possible images in a high-dimensional manifold is virtually infinite, "vastly…complicated and impossible to fully characterise".[333] Regardless of what is learnt, an attacker will be able to generate more (unexpected) adversarial distortions, with a large subset of these slipping through the net and spoofing the system.[334]

---

[327] Cited in ibid., 184.

[328] See JASON (n 325), 30, for further elaboration, using the panda/gibbon example.

[329] Szegedy et al. (n 315); Nguyen, Yosinski and Clune (n 321); Goodfellow, Shlens and Szegedy (n 324).

[330] Ian Goodfellow, *Presentation at Re-Work Deep Learning Summit* (24 February 2015) <https://www.youtube.com/watch?v=Pq4A2mPCB0Y> accessed 16 May 2018.

[331] Goodfellow, Shlens and Szegedy (n 324), 5 (reporting a fall in the error rate on classifying false images, from 89% before adversarial training, to 18% after training).

[332] Email from Ian Goodfellow to Maziar Homayounnejad (18 May 2018), on file with author.

[333] JASON (n 325), 31.

[334] Nguyen, Yosinski and Clune (n 321).

2.5.7 Broader Implications for the Military Use of AI

Aside from image classification, this issue reminds us more broadly that learning machines are essentially 'black boxes', which exhibit "counterintuitive and unexpected forms of brittleness",[335] and this makes it very difficult for commanders – indeed, even software engineers – to accurately predict the circumstances in which a network might fail.[336] As Szegedy et al. put it, the current class of DNNs have "nonintuitive characteristics and intrinsic blind spots, whose structure is connected to the data distribution in a non-obvious way";[337] namely, the networks are inherently *inscrutable* and *non-intuitive*. The JASON group of scientific experts, which recently reported on the implications of AI for the US DoD, similarly concluded:

> [T]he sheer magnitude, millions or billions of parameters (i.e. weights/biases/etc.), which are learned as part of the training of the net…makes it impossible to really understand exactly how the network does what it does. Thus the response of the network to all possible inputs is unknowable.[338]

Consequently, Clune argues that DNNs for lethal autonomous targeting may be unacceptably vulnerable to adversarial hacking;[339] not only to conceal legitimate targets but also to draw lethal firepower towards protected persons and objects,[340] perhaps in a propaganda war.

However, the solution is not necessarily to ban the use of AI, or even a more narrowly-tailored ban on DNNs in military applications. As the JASON group noted, despite being inherently prone to both error and adversarial attack, DNNs may still be incorporated as components within a larger/hybrid system, where other pieces of the system have a "supervisory role".[341] This may enable robust validation and verification by proxy, in relation to the entire system.[342]

---

[335] Scharre (n 13), 182.

[336] Ibid.

[337] Szegedy et al. (n 315), 2.

[338] JASON (n 325), 28-29.

[339] Scharre (n 13), 187.

[340] Ibid.

[341] JASON (n 325), 32.

[342] Ibid. See also 36-37 for current examples of hybrid architectures.

2.6 Conclusion

This chapter has proposed a working definition of LAWS that incorporates four relevant dimensions: human-machine interaction; the task being delegated; machine complexity; and complexity of the operational environment. It has also delineated the relevant near-term LAWS as wide-area loitering systems, operating standalone or in swarms. Given the risk constraint on development and deployment, it is likely that such systems will initially be fielded in the maritime area, followed by aerial and ground-based systems; first for anti-material targeting, then expanding into anti-personnel targeting and more varied attack missions as the technology, the Concepts of Operations, and the Tactics, Techniques and Procedures all develop.

Crucially, the chapter has also demonstrated an inherent paradox: while LAWS will be sophisticated machines employing various forms of advanced AI that may *appear* to replicate human thought processes, the systems will essentially be executing *technical* processes, with no more 'understanding' of their actions than the landmine depicted in Figure 2.1. This is epitomised by the fact that autonomous systems will act as 'optimisers', potentially resulting in "one objective being pursued relentlessly despite other common-sense values being salient";[343] again, not too different from the landmine responding only to a P > X constraint. In such instances, perverse results are a real risk, unless humans carefully program and meaningfully supervise the systems. There are numerous other sources of brittleness that also suggest the need for human control in autonomy, the most vivid being the adversarial examples that so easily spoof otherwise sophisticated computer vision systems. These difficulties, which arise from the inscrutable and non-intuitive nature of machine learning, are in addition to limitations in sensing and GPS/communications technologies. Accordingly, LAWS and its associated AI will be rather brittle by nature.

Even if this characteristic can be mitigated over time with better components, programming and testing and evaluation, it still indicates a fundamental difference between human and machine cognition. Hence, the International Panel on the Regulation of Autonomous Weapons (iPRAW) considers that the terms AI and machine learning inappropriately anthropomorphise these safety-critical systems, and may lull people into a) a false sense of security regarding their capabilities, and b) a

[343] UNIDIR (n 87), 11.

false impression of 'intention and purpose'.[344] Instead, iPRAW recommends the terms 'computational methods' (in place of AI) and 'evolving algorithms' or 'data-driven algorithms' (in place of machine learning).[345] While this thesis will continue to use the traditional terms for their familiarity, it nonetheless acknowledges the validity of the reasoning that led to iPRAW's preferred terminology.

Yet, this does not necessarily mean that LAWS should be banned. As the JASON group noted, above, hybrid systems may provide a solution, and this includes systems integrating a human operator who can bring deliberative thinking into the process. Accordingly, Chapter 3 will consider the broad legal implications of weapons autonomy and Chapter 4 will examine the 'meaningful human control' concept at the high level, and as a precursor for the more detailed application of the LOAC rules that will follow in subsequent chapters.

[344] iPRAW (n 214), 13.
[345] Ibid., 9-10.

Chapter 3
Broad Legal Implications of Weapons Autonomy

3.1 Introduction

The following chapter contains two main sections, which will examine in broad terms the legal implications of autonomous attack. First, 3.2 builds on Chapter 2 to argue that lethal autonomous weapon systems (LAWS) will be mere tools: both technical in nature, and devoid of any of the human qualities required to have legal obligations, or to exercise broad legal judgment. Hence, they will not 'apply the law' as such, so academic analyses and State parties at the current diplomatic process in Geneva err when they query whether LAWS can 'comply with IHL'. International humanitarian law (IHL) obligations are addressed *exclusively* to humans; namely, engineers, programmers, commanders and weapons operators (WOs), who will have to ensure legal compliance in their design, development, deployment and use of LAWS. Accordingly, the relevant question is whether LAWS can be designed, developed, deployed and used *in compliance with* IHL, or the law of armed conflict (LOAC). Second, 3.3 sketches the contours of how we might expect the fielding of LAWS to affect the legal analysis of targeting. Here, it will be seen that autonomy will *reassign* some operational decisions between human actors and it will necessitate their *earlier timing*, thus making them more *general* in character. This may be expected to weaken the causal nexus between deliberative human decisions, and specific actions and outcomes on the battlefield. Consequently, there will be a need for appropriate and meaningful human control, to mitigate the harshness of full autonomy – an issue that will be addressed in Chapter 4.

3.2 LAWS Will Be Mere Tools, Not Persons

There has long been a tendency to anthropomorphise relatively simple devices and systems, which in reality have very little in common with human beings.[1] This is even more common with computers[2] and, arguably, can be expected to a far greater extent

[1] Nick Bostrom, 'The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents' (2012) 22 Minds & Machines 71, 72-73 (discussing the naïve observer who attributes human-like qualities to cars and vending machines).

[2] Batya Friedman and Peter H. Kahn, 'Human Agency and Responsible Computing: Implications for Computer System Design' in Batya Friedman (ed.), *Human Values and the Design of Computer Technology* (CSLI, 1997), 221 (presenting evidence that even educated and regular computer users tend to assign agency and hold computers partially responsible for computer errors).

with robotic devices – particularly those designed to interact with humans.[3] On the battlefield, soldiers are reported to have anthropomorphised their tele-operated *PackBot* Explosive Ordnance Disposal (EOD) robots, often attributing personality quirks to the machines, giving them names and incorporating them into their units.[4] Furthermore, when *PackBots* are destroyed during a military operation, soldiers have shown visible signs of grief and a sense of loss,[5] in some cases insisting on a military 'funeral' for their 'fallen comrade'.[6]

Such anthropomorphisms are very common: indeed, some have argued that "humans are wired to anthropomorphize",[7] and that this will frequently shape their interactions with machines.[8] For example, whether such devices are regarded as a 'friend' or a 'part of the family' in a civilian context; or as a 'comrade' or, potentially, a 'war criminal' in a military context.[9] Yet, this tendency to humanise inanimate objects can have catastrophic effects in safety-critical systems, especially when combined with the inherent brittleness of artificial intelligence (AI) seen in Chapter 2. Along with the risk of 'automation bias',[10] and the distance and psychological detachment that often occur in unmanned systems, the result may be an apparent reassignment of moral agency to machines, thereby creating a "moral buffer" and a *perception* amongst WOs that they are not accountable for their decisions.[11]

For legal purposes, however, LAWS are decidedly machines, not persons; only the latter can be accountable and held responsible for battlefield decisions. Despite academic analyses that typically query whether LAWS will be able to 'comply with

---

[3] Robert Sparrow and Linda Sparrow, 'In the Hands of Machines? The Future of Aged Care' (2006) 16 Minds & Machines 141, 155 (focusing specifically on elder-care robots).

[4] Paul J. Springer, *Military Robots and Drones: A Reference Handbook* (ABC-CLIO, 2013), 186-89.

[5] Peter W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (Penguin, 2009), 337-38.

[6] Springer (n 4), 189.

[7] Neil M. Richards and William D. Smart, 'How Should the Law Think About Robots?' in Ryan Calo, A. Michael Froomkin and Ian Kerr (eds.), *Robot Law* (Edward Elgar, 2016), 20.

[8] Ibid.; Kate Darling, '"Who's Johnny?" Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy' in Patrick Lin, Keith Abney and Ryan Jenkins (eds.), *Robot Ethics 2.0* (OUP, 2017).

[9] Ibid.

[10] This is where human decision-makers over-trust automated systems and either disregard, or do not notice erroneous/contradictory information, hence uncritically implementing computer-generated recommendations.

[11] ML. Cummings, 'Creating Moral Buffers in Weapon Control Interface Design', *IEEE Technology and Society Magazine* (Fall 2004) 28, 31-33.

IHL',[12] and the many calls by States parties in Geneva to ensure this,[13] the real question is whether LAWS can be designed, developed, deployed and used *in compliance with* IHL/LOAC. This legal reality is based on two distinct but related strands of argument: the technical nature of LAWS, and the absence of human qualities.

3.2.1 The Technical Nature of LAWS

First, as was seen in Chapter 2, even the most advanced weapon systems are merely inanimate devices that follow technical processes based on the 'sense-think-act' paradigm;[14] there is no 'free will' in the philosophical sense. This will remain true even as advances in AI and machine learning bring ever-greater levels of sophistication to weapon systems. As mentioned in 2.2.2, and especially in relation to Figure 2.3, LAWS will entail software-based controllers 'stepping into the shoes' of the human soldier, both to operate the weapon and to monitor the target, the environment and the weapon system itself. Yet, these controllers are merely special-purpose computers running on the 'stored program' concept.[15] Namely, they are 'calculating machines' that store instructions (entered by a human programmer) and data in the same internal memory unit.[16] Subsequently, both are processed together by the central processing unit's arithmetic sub-unit, so that in the course of a computation, the instructions are not just executed but also modified at electronic speeds, to effectively govern the controller's operation.[17] Assuming no major technological shift on the horizon, autonomous systems will employ essentially the same technology: controllers run by software, which in turn is written by human programmers. Accordingly, Richards and Smart note that:

> Robots are, and for many years will remain, tools. They are sophisticated tools that use complex software, to be sure, but no different in essence than…a word processor, a web browser, or the braking system in your car.[18]

---

[12] For example, Robin Geiss, *The International Law Dimension of Autonomous Weapons Systems* (Friedrich Ebert Stiftung Study, October 2015), 13-17.

[13] See the various statements and working papers of States parties at the 2018 Group of Governmental Experts (GGE) Meeting on Lethal Autonomous Weapons Systems (LAWS) <https://www.unog.ch/80256EE600585943/(httpPages)/7C335E71DFCB29D1C1258243003E8724?OpenDocument> accessed 21 May 2018.

[14] Tim McFarland, 'Factors Shaping the Legal Implications of Increasingly Autonomous Military Systems', (2015) 900 International Review of the Red Cross 1313.

[15] William Aspray, 'Back to Basics: The Stored Program Concept' (1990) 27 IEEE Spectrum 51.

[16] Ibid.

[17] Ibid.

[18] Richards and Smart (n 7), 18.

This is axiomatic of the simpler rules-based systems, where it may be easy to see that the machine is not acting independently in any legally significant way. However, even with the more complex future LAWS, which will be goal-directed and will exercise 'discretion' and 'self-direction', the programs running such machines are still just sets of pre-defined instructions.[19] Consequently, as discussed in 2.3.4, an identical set of inputs (programming, operator commands and battlefield conditions) should lead to the same output (selection and engagement of a specific target) every time. Yet with subtle variations in those inputs sometimes leading to large differences in the 'choices' being made by the machine, observers may perceive different behaviours in apparently identical situations, which in turn may be seen as 'free will' on the part of the machine.[20] Arguably, this would be an erroneous way to think about a weapon system, which will always remain a tool of the individual commander or operator.[21] Again, Richards and Smart capture the essence of the argument, by asserting:

> As the autonomy of the system increases, it becomes harder and harder to form the connection between the inputs (your commands) and the outputs (the robot's behavior), but it exists and is deterministic.[22]

A fortiori, there may be a strong perception of 'free will' in the case of a machine learning LAWS, which may alter part of its own algorithm in relation to its critical functions, thereby actually leading to different system behaviours at the low-level. Even so, there is arguably still no reason to see this as any more than an exercise in software development – one that is controlled by humans, similar to the simpler and more clearly deterministic programming. In both instances, the developer defines some desired behaviour for the system and writes a program designed to impart that behaviour to the machine.[23] The only real distinction is in the algorithm employed: instead of directly encoding actions to be taken, as would the developer of a relatively simple program, the designer of a learning machine specifies desired outputs, activation functions and learning functions intended to formulate an optimum set of

---

[19] McFarland (n 14), 14-15.

[20] Richards and Smart (n 7), 18.

[21] William Boothby, 'Dehumanization: Is There a Legal Problem Under Article 36?' in Wolff Heintschel von Heinegg, Robert Frau and Tassilo Singer (eds.), *Dehumanization of Warfare: Legal Implications of New Weapons Technologies* (Springer, 2018).

[22] Richards and Smart (n 7), 18.

[23] McFarland (n 14), 15.

actions in response to environmental stimuli.[24] Accordingly, there is an extra layer of abstraction between the developer and the selection and engagement of the target.[25] This may obscure the process of matching specific (attack) outcomes to specific (human developer) commands, but it does not change the fact that even a machine learning LAWS will only be executing (at the broad level) *ex ante* instructions formulated by its developer.[26]

The above references to computing technology are included purely to underscore the fact that LAWS – even the most sophisticated ones – will likely follow a deterministic path, based on the physical manipulation of symbols (digital data). This should counsel against any tendency to be misled by manufacturer references to 'choice' or 'truly autonomous' features[27] – terms which may hold marketing sway, but arguably no real technical or legal significance. It should also discourage any notion that LAWS may become an 'intermediate category' between weapon systems and combatants.[28] To reiterate: computers do not choose whether or not to run a program stored in memory; nor do they decide whether or not to execute a particular instruction within a program.[29] Any appearance of 'choice' is the result of an intricate web of instructions within a complex software code and a cluttered environment. The only function of a computer – LAWS included – is to run whatever software is installed on it.

3.2.2 The Absence of Necessary Human Qualities

The second strand of the argument relates to the common assumption that LAWS will be neither sentient nor self-aware,[30] and they will not be able to exercise any metacognition.[31] To an extent, this may be inferred from the first strand, in that being limited to technical capacities *a priori* means LAWS cannot possess any of the human qualities noted above. Yet, it is worth expanding on these points for their

[24] See 2.5.1.2-2.5.1.3 on machine learning and its methods.

[25] McFarland (n 14), 15.

[26] Ibid.

[27] Gary E. Marchant et al., 'International Governance of Autonomous Military Robots' (2011) 12 Columbia Science and Technology Law Review 272.

[28] Cf. Hin-Yan Liu, 'Categorization and Legality of Autonomous and Remote Weapons Systems' (2012) 94 International Review of the Red Cross 627.

[29] McFarland (n 14), 15.

[30] For example, Boothby (n 21).

[31] Eliav Lieblich and Eyal Benvenisti, 'The Obligation to Exercise Discretion in Warfare: Why Autonomous Weapons Systems are Unlawful' in Nehal Bhuta et al. (eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (CUP, 2016), 250.

contradistinguishing features and as further justification of the need for human control in autonomy.

3.2.2.1 LAWS Will Be Neither Sentient nor Self-Aware

3.2.2.1.1 Sentience

Sentience refers to "the capacity to experience pleasure or pain".[32] It is a prerequisite for having subjective interests;[33] and, therefore, is a ground for moral status,[34] which in turn gives rise to moral obligations.[35] This leads to a number of legal protections afforded to the entity enjoying moral status.[36] While LAWS will certainly process sensory information, which is a prerequisite for sentience, this will only be to determine such activities as navigation; and the tracking, prioritisation, selection and engagement of targets. It is assumed there will be no capacity to feel pleasure or pain as such,[37] and this highlights the illogicality of soldiers anthropomorphising their *PackBots*.

3.2.2.1.2 Self-Awareness

Self-awareness (SA) goes significantly beyond sentience, and broadly refers to the capacity for introspection. On some views, SA is neither static nor singular, but is a dynamic process, the most advanced stage of which is where "[t]he self is now recognized not only from a first person perspective, but also from a third person's".[38] Accordingly, self-aware entities "are not only aware of what they are, but *how* they are in the minds of others: How they present themselves to the public eye".[39] This is a prerequisite for the *mens rea* element of criminal liability, where this requires specific intent.[40]

---

[32] Agnieszka Jaworska and Julie Tannenbaum, 'The Grounds of Moral Status' in Edward N. Zalta (ed.) *Stanford Encyclopedia of Philosophy* (Spring 2018) <https://plato.stanford.edu/archives/spr2018/entries/grounds-moral-status/> accessed 21 May 2018.

[33] To have 'interests' depends on being "capable of suffering and enjoyment" or on having "desires, preferences, or concerns": David DeGrazia, *Taking Animals Seriously: Mental Life and Moral Status* (CUP, 1996), 40.

[34] Peter Singer, *Practical Ethics* (2nd ed., CUP, 1993).

[35] James G. Dwyer, *Moral Status and Human Life* (CUP, 2010), 9.

[36] Ibid., Chapter 1.

[37] Mark Bishop, 'Why Computers Can't Feel Pain' (2009) 19 Minds and Machine 519.

[38] Philippe Rochat, 'Five Levels of Self-Awareness as They Unfold in Early Life' (2003) 12 Consciousness and Cognition 717, 722.

[39] Ibid.

[40] John Buyers, 'Liability Issues in Autonomous and Semi-Autonomous Systems', *Osborne Clarke Publication No. 0000000* (2015), 6 <http://www.osborneclarke.com/media/filer_public/c9/73/c973bc5c-cef0-4e45-8554-

Both sentience and SA (SSA) are distinctly human and animal qualities, with which it is assumed no artificial machine will be endowed,[41] at least in the foreseeable future.[42] This is different to the question of whether machines can perform at human or super-human levels of visual intelligence, or whether they can sense human emotional states.[43] Such capacities are merely the result of formal computational tasks, and are qualitatively different to SSA.

3.2.2.1.3 Legal Consequences

Yet SSA are necessary characteristics for an entity to be recognised as a (natural) legal person capable of personal responsibility. As Solum points out, humans are legally recognised as 'persons' because of their intuitions and shared experiences.[44] Matambanadzo builds on this, focusing on the embodied human being and arguing that such embodiment allows a legal entity to "draw[] on shared intuitions about who counts in our community of legal persons and how we should take account of them".[45] Thus, while being human is not necessary for 'personhood', the average adult human has the capacity to exercise rights and to owe duties, and this is a strong driver towards ascribing legal personhood.[46] However, without the common human traits of SSA and autonoetic metacognition (see below) it is impossible to ascribe personhood for accountability and responsibility to a LAWS. Thus, Sassóli asserts "[t]he difference between a weapon system and a human being is not quantitative but qualitative; the two are not situated on a sliding scale, but on different levels – subjects and objects".[47]

---

f6f90f396256/itech_law.pdf> accessed 21 May 2018. See also Deborah W. Denno, 'A Mind to Blame: New Views on Involuntary Acts' (2003) 21 Behavioural Sciences and the Law 601, 611 (discussing the absence or impairment of self-awareness as a prerequisite for a finding of involuntariness, thereby negating *mens rea*).

[41] Richard H. Schlagel, 'Why Not Artificial Consciousness or Thought?' (1999) 9 Minds and Machines 3.

[42] Riccardo Manzotti, 'The Computational Stance is Unfit for Consciousness' (2012) 4 International Journal of Machine Consciousness 401.

[43] See 2.5.1-2.5.2 on artificial intelligence and deep learning.

[44] Lawrence B. Solum, 'Legal Personhood for the Artificial Intelligences' (1992) 70 North Carolina Law Review 1231, 1285.

[45] Saru M. Matambanadzo, 'The Body, Incorporated' (2013) 87 Tulane Law Review 1, 50.

[46] Alexis Dyschkant, 'Legal Personhood: How We Are Getting it Wrong' (2015) University of Illinois Law Review 2075, 2080.

[47] Marco Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified' (2014) 90 International Law Studies 308, 323.

More specifically and relevant to LAWS – the actions of which may potentially violate IHL/LOAC and international criminal law (ICL) – is the question of personhood for criminal liability.[48] In this connexion, it is worth noting that there are four general purposes of punishment for a criminal offence: retribution, deterrence, rehabilitation and incapacitation.[49] While it is beyond the scope of this thesis to delve into these, a few brief reflections on the first two are worth making. First, retribution arguably presumes the ability to feel pain and suffering; without sentience, however, this is negated in a LAWS. Second, deterrence is also arguably negated, given the automatic and non-discretionary running by a LAWS of whichever software is installed on it, free from any self-awareness or fear. Clearly, without the capacity to make conscious choices, there can be no deterrent effect of any criminal laws or threat of punishment.

Thus, it is highly unlikely that in the near-term LAWS will acquire a separate legal identity capable of being held accountable for violations of IHL or ICL. Arguably, no AI system will possess the kinds of capacities that would justify such a jurisprudential change *in the LOAC*.[50] Of course, this sort of change may be possible – even desirable from a policy perspective – in civilian fields such as manufacturing and commercial services: there, personhood will enable robotic systems to be identified with distinct revenue streams, much like corporations are; and possibly for similar reasons. But such reasoning arguably does not extend to battlefield robots, which are decidedly objects not subjects. As Sassóli succinctly puts it: "[a] combatant is a human being, only he or she is an addressee of legal obligations".[51]

To conclude, it is necessary to guard against anthropomorphising LAWS. As Richards and Smart have argued, albeit in the context of humanoid robots, failure to heed this warning may mean that "we might hold the designers less responsible for [the robot's] actions",[52] because if "it seems to have some limited form of free will…how can we expect the designers to cover every eventuality?"

---

[48] See, more generally, Mohamed Ellewa Badar, *The Concept of Mens Rea in International Criminal Law: The Case for a Unified Approach* (Hart Publishing, 2015).

[49] Gabriel Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (Springer International, 2015), 185.

[50] Cf. Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate, 2009), 105.

[51] Sassóli (n 47), 323.

[52] Richards and Smart (n 7), 18.

> We must avoid the Android Fallacy. Robots, even sophisticated ones, are just machines. They will be no more than machines for the foreseeable future, and we should design our legislation accordingly. Falling into the trap of anthropomorphism will lead to contradictory situations…[53]

Arguably, this reasoning applies more strongly in the case of LAWS, where battlefield errors can have particularly catastrophic consequences, yet would not be adequately dealt with by assuming machine personhood, or any form of intermediate category. Some have argued that this will leave a lacuna in the law, by way of an accountability and responsibility (A&R) gap,[54] and that this justifies a pre-emptive ban.[55] However, there is a strong body of academic opinion that rebuts such a claim. Some argue that the legality of a weapon system has never hinged on issues of personal accountability,[56] instead focusing on the liability of weapons designers and procurement teams (in developing and fielding the systems);[57] or on commanders and WOs (in deploying and using LAWS on the battlefield).[58] Others argue that A&R problems are rather isolated and have practical workaround solutions.[59] Alternatively, where current law is seen not to be fit for purpose, some have argued for a parallel legal regime to plug the gap via the law of torts.[60] Consequently, even if LAWS do give rise to some A&R challenges, these are likely to be "far smaller than some critics of military robots believe".[61] As will be seen in Chapter 4, so long as there is meaningful human control during the design, development, deployment and use of LAWS, there is arguably no reason to believe that an unavoidable A&R gap will take hold.

---

[53] Ibid., 20.

[54] For example, Human Rights Watch, *Mind the Gap: The Lack of Accountability for Killer Robots* (Human Rights Watch, 2015); Markus Wagner, 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems' (2014) 47 Vanderbilt Journal of Transnational Law 1371, 1399-1409.

[55] Human Rights Watch, ibid.

[56] Charles J. Dunlap Jr., 'Accountability and Autonomous Weapons: Much Ado About nothing?' (2016) 30 Temple International & Comparative Law Journal 63, 65-66.

[57] Geoffrey S. Corn, 'Autonomous Weapons Systems: Managing the Inevitability of 'Taking the Man Out of the Loop'' in Bhuta et al. (eds.) (n 31).

[58] Dunlap (n 56), 68-73; Michael N. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) Harvard National Security Journal Features 1, 33.

[59] For example, Sassóli (n 47), 325 (arguing that the restrictive temporal field of application in ICL can be addressed by treating programmers of a 'war crime algorithm' as indirect perpetrators, or as guarantors obliged to intervene once armed conflict begin).

[60] Rebecca Crootof, 'War Torts: Accountability for Autonomous Weapons' (2016) 164 University of Pennsylvania Law Review 1347.

[61] Krishnan (n 50), 105.

3.2.2.2 Might LAWS Have the Capacity for Metacognition?

Metacognition refers to "…one's knowledge concerning one's own cognitive processes and products or anything related to them".[62] At its simplest, this refers to the higher-order process of "thinking about thinking"[63] or "knowing about knowing".[64] At a more advanced level, metacognition enables key critical thinking skills, and it comprises two distinct components: *knowledge* about cognition; and *regulation* and *orchestration* of cognition.[65] Thus, metacognition takes on an 'executive role', which enables humans to a) diagnose their current state of knowledge, and b) to actively control their learning, to acquire further skills and knowledge in a relatively more efficient way.[66] As will be seen below, these competences are key to the effective application of LOAC norms.

Concretely, a metacognitive entity knows what it knows and what it does not know; but beyond this, it has the potential, through education, training, and personal experience, to attain a variety of key evaluative skills.[67] For example, it can be trained to:

- know how and why it arrived at a particular answer;
- exploit means and develop ways to acquire missing information needed to get to an answer, or to perform a task;
- interrogate the validity and reliability of sources, and prioritise between them.
- connect newly-gathered information to its existing knowledge base and personal experience, and to adapt each in light of the other, with appropriate weighting given to each one; and
- make generalisations and analogies, to transfer knowledge from one instance or subject area to another.

---

[62] John H. Flavell, 'Metacognitive Aspects of Problem Solving' in Lauren B. Resnick (ed.), *The Nature of Intelligence* (Erlbaum Associates, 1976), 232.

[63] Deanna Kuhn and David Dean, 'Metacognition: A Bridge Between Cognitive Psychology and Educational Practice' (2004) 43 Theory into Practice 268, 270.

[64] Ruth Garner and Patricia Alexander, 'Metacognition: Answered and Unanswered Questions' (1989) 24 Educational Psychologist 143.

[65] Gregory Schraw, 'Promoting General Metacognitive Awareness' (1998) 26 Instructional Science 113, 114.

[66] Accordingly, it is one of the hallmarks of general intelligence. See 2.5.1 on narrow *versus* general AI.

[67] "I am engaging in metacognition if I notice that I am having more trouble learning *A* than *B*; if it strikes me that I should double-check *C* before accepting it as a fact;…if I become aware that I am not sure what the experimenter really wants me to do; if I sense I had better make a note of *D* because I may forget it; if I think to ask someone about *E* to see if I have it right": Flavell (n 62), 232.

These cognitive actions must meet two structural preconditions: agents must be able to assess whether the task considered is within their reach and solvable within the given time span; and once the action is performed, they must be able to reliably evaluate its success or failure.[68] Moreover, the extent to which an agent is willing to expend time and effort over such actions, and the threshold of confidence sufficient for triggering an overt action, will often be influenced by perceptions of how critical the main task is.[69] As LOAC-based decisions often occur in a strongly safety-critical context, the extent of metacognitive probing by commanders, their battle staffs, and individual soldiers in the field can be expected to be relatively high.[70]

3.2.2.2.1 A Distinctly Human Trait, Necessary for Applying the LOAC

At first sight, these higher-order skills appear to be distinctly human traits, and this is underscored by much of the literature being in the area of educational psychology. Crucially, a number of key LOAC obligations, as articulated in Additional Protocol I[71] (AP I), would seem to presuppose metacognitive thinking. For example, the requirement to take "constant care" to spare civilian lives and property;[72] to give 'effective advance warning' of attacks "unless circumstances do not permit";[73] the pervasive requirement for "feasibility" in the circumstances;[74] or, indeed, the obligation to presume civilian status in situations of 'doubt'.[75] Arguably, the most 'cerebral' and abstract judgment that calls for metacognitive thinking is the principle of proportionality: this prohibits attacks in which the estimated 'collateral damage' is "excessive" in relation to the concrete and direct military advantage anticipated,[76] and is considered to be a profoundly "human qualitative…decision".[77]

---

[68] Joëlle Proust, *The Philosophy of Metacognition: Mental Agency and Self-Awareness* (OUP, 2013).

[69] Ibid.

[70] Though it is acknowledged that with a battlefield soldier facing an imminent mortal threat, instinct and urgency will often substitute metacognitive probing.

[71] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3.

[72] Article 57(1), AP I.

[73] Article 57(2)(c), AP I.

[74] Article 57(2)(a)(i) and (ii), AP I.

[75] Articles 50(1) and 52(3), AP I, in relation to persons and objects, respectively.

[76] Articles 51(5)(b) and 57(2)(b), AP I.

[77] Noel E. Sharkey, 'The Evitability of Autonomous Robot Warfare' (2012) 94 International Review of the Red Cross 787, 789-90.

Common to many of these obligations is that they are framed as 'principles' or 'standards', rather than 'rules'[78] (as distinct from 'legal rule' in the broader sense of referring to all legal norms). The difference between the specific categories hinges not on their normative status, which in all cases is equally binding, but on whether the precise content of the norm is determined *before* (rules) or *after* (standards/principles) relevant facts have materialised.[79] Thus, standards/principles are written in relatively vague language, so when applying them to specific facts difficult value judgments are inevitable.[80] By contrast, rules are drafted more precisely and the process of matching these with concrete facts is more technical,[81] hence more amenable to *ex ante* programming and application to machine-perceptible facts.

Giving content to a norm requires effort to "analyze a problem [and] resolve value conflicts",[82] very much in line with the metacognitive traits bullet-pointed above. Yet, resolving value conflicts is almost impossible for a machine to do, save for a rapid mathematical solution which humans will be unlikely to identify as 'sound judgment'.[83] Machines apply rules or procedures, which must "specify every element in sufficient detail for a computer to be able to operate on it".[84] In this connexion, Asaro contrasts chess with LOAC: AI systems have long outperformed humans at chess, because it is "a fairly well-defined rule-based game that is susceptible to computational analysis".[85] The same applies to the game of Go, which is vastly more

---

[78] Louis Kaplow, 'Rules Versus Standards: An Economic Analysis' (1992) 42 Duke Law Journal 557.

[79] Ibid., 568-86 (discussing the notion of "*ex ante* versus *ex post* creation of the law").

[80] While standards and principles are used interchangeably here, it is acknowledged that some scholars see a distinction between the two, depending not on their normative status for those who have to *follow* them, but on the "extent to which they constrain those who are charged with *applying* them". See Lawrence Solum, 'Legal Theory Lexicon: Rules, Standards and Principles', *Legal Theory Blog* (6 September 2009) (emphasis added) <http://lsolum.typepad.com/legaltheory/2009/09/legal-theory-lexicon-rules-standards-and-principles.html> accessed 18 September 2018.

[81] Ibid. (summarising the distinction in that a **rule** is "cast in the form of a bright-line"; a **standard** is usually "in the form of a balancing test" with an "exhaustive set of considerations for adjudication"; while a **principle** provides "mandatory [but non-exhaustive] considerations for judges" and it acts as "guidance for the interpretation or application of a rule or standard").

[82] Kaplow (n 78), 621.

[83] See the hypothetical example on minimising civilian casualties with reverberating effects in Chapter 2 (notes and text accompanying) nn 244-245. See also UNIDIR, 'The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches', *UNIDIR Resources, No. 6* (2017), 11 <http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf> accessed 21 May 2018.

[84] Sharkey (n 77), 789. See also 2.2.3.2 on the need for *precision* and *tangibility* in task execution.

[85] Peter Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making' (2012) 94 International Review of the Red Cross 687, 705.

complex but still underpinned by precise rules and tangible outcomes in relation to which AI systems have outperformed humans.[86] By contrast, LOAC requires "a great deal of interpretive judgment to be applied appropriately in any given situation", and it is to a great extent "a matter of social norms",[87] "situational awareness" and "the use of common sense";[88] all of which will very likely elude a software algorithm.[89] This is because the "strategic vagueness" of broad standards and principles presupposes certain cognitive capabilities, which are required to map the conditions assumed by those standard to concrete situations;[90] yet, most of these capabilities cannot be specified in machine-executable code.[91] Asaro also points to the stability of rules in chess *versus* the shifting and competing/conflicting interpretations of LOAC from day to day, and even *within* the same day inside the same conflict.[92] The author concludes that it is incumbent upon persons applying lethal force to take all these perspectives into account and to draw insight from them, before making life and death decisions.[93] This is necessary in order to question the LOAC standards and the appropriateness of their application in a given factual scenario, thereby minimising error in perilous and irreversible situations.[94] As noted above, such a duty to consider and potentially reconcile shifting/conflicting viewpoints arguably calls for the most probing of metacognitive thinking: to assess the state of knowledge in the 'fog of war',[95] understand where there are gaps preventing a reasoned conclusion, and either to locate the required information and integrate it into the overall assessment (time permitting);

---

[86] Dan Silver et al., 'Mastering the Game of Go Without Human Knowledge' (2017) 550 Nature 354.

[87] Asaro (n 85), 699, 705.

[88] Sharkey (n 77), 789.

[89] Unlike, for example, some areas of road traffic law, which remain rules-based and strict liability, with the result that legal advice has been successfully automated. See, for example, Samuel Gibbs, 'Chatbot Lawyer Overturns 160,000 Parking Tickets in London and New York', *The Guardian* (28 June 2016) <https://www.theguardian.com/technology/2016/jun/28/chatbot-ai-lawyer-donotpay-parking-tickets-london-new-york> accessed 21 May 2018.

[90] Lucy Suchman, 'Situational Awareness and Adherence to the Principle of Distinction as a Necessary Condition for Lawful Autonomy' in Robin Geiß (ed.), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016), 279 (referring specifically to 'situational awareness', and arguing that when such cognitive capabilities are in place, the openness of standards is – far from being a problem – exactly what makes them effective as a reference point for concrete action).

[91] Ibid. (pointing out that no progress has been made in AI towards operationalising 'situational awareness' for indeterminate environments).

[92] Asaro (n 85), 699.

[93] Ibid., 700-03.

[94] Ibid.

[95] Carl von Clausewitz (author), FN. Maude (ed.), *On War*, Book I, Chapter III (Wordsworth Classics, 1997), 42 ("War is the province of uncertainty: three-fourths of those things upon which action in war must be calculated are hidden more or less in the clouds of great uncertainty").

or to go ahead with a 'satisficing' decision in the circumstances (where time-pressured).[96] Such judgment calls presuppose not only metacognitive thinking but, specifically, one that is informed by human instincts as developed through experience.[97]

This prevalence of standards over rules in the LOAC is not just a recent phenomenon but also looks set to rise in the foreseeable future,[98] thereby posing greater difficulties for any putative application of LOAC by machines. Accordingly, Lieblich and Benvenisti argue that LAWS will – almost by definition – be unable to discharge numerous legal obligations during the conduct of hostilities, as these require battlefield actors to exercise *unfettered* discretion, which in turn requires metacognitive thinking.[99] In contradistinction, LAWS embody the *pre-bound* discretion of their programmers and deploying commanders. Such discretion will be operationalised in precise rules that will often preclude 'fresh thinking' or experiential insight as new and specific circumstances arise.[100]

3.2.2.2.2 Is There an Emerging Machine Metacognition?

Notwithstanding, the concept of metacognition has been making its way into the computer science literature,[101] and has even been explicitly applied in machine learning. For example, Babu and Suresh present an algorithm for a 'metacognitive neural network' (McNN) classifier,[102] which consists of a *cognitive* component and a *metacognitive* component.[103] The latter adapts learning strategies by deciding *what*, *when* and *how* to learn,[104] with the result that the McNN classifier exhibits superior

---

[96] Herbert A. Simon, *Models of Man: Social and Rational* (Wiley, 1957), 204-05.

[97] Dan Saxon, 'What is 'Judgment' in the Context of the Design and Use of Autonomous Weapon Systems?' in Geiß (ed.) (n 90).

[98] Amichai Cohen, 'Rules and Standards in the Application of International Humanitarian Law' (2008) 41 Israel Law Review 41 (noting the growing acceptance of international courts applying IHL/LOAC, which lends itself to the standards approach and decisions being taken *ex post*. In turn, more interpretive judgment may be expected from battlefield commanders and their legal advisers).

[99] Lieblich and Benvenisti (n 31) (assuming war to be a form of governance, and applying a global administrative law approach to argue for a legal requirement for 'unfettered discretion' on the battlefield, which can only be carried out by fully metacognitive persons. Hence, the authors argue that LAWS are *per se* unlawful by reason of the pre-bound discretion the machines will necessitate).

[100] Ibid.

[101] For a general overview, see Michael T. Cox, 'Metacognition in Computation: A Selected Research Review' (2005) 169 Artificial Intelligence 104.

[102] G. Sateesh Babu and Sundaram Suresh, 'Meta-Cognitive Neural Network for Classification Problems in a Sequential Learning Framework' (2012) 81 Neurocomputing 86.

[103] Ibid., 88.

[104] Ibid., 88, 89.

performance when compared with standard machine learning classifiers.[105] Subsequently, this same metacognitive approach was applied to 'extreme learning machines',[106] resulting in a 'metacognitive extreme learning machine' (McELM) that outperformed existing McNN classifiers, with less computational effort.[107] In the broader field of robotics, metacognition has been applied to the concept of the 'robot baby' in a room. This starts life without a robust self-model, but has the primary goal of learning about itself and exploring its environment.[108] In turn, a Metacognitive Loop (MCL) autonomously guides the creation of sub-goals and plans that contribute to achieving the primary goal.[109] The MCL also coordinates the various types of learning (supervised, unsupervised or reinforcement) that the robot will undertake: both internal (learning about its own state and processes); and external (learning about its environment).[110]

Accordingly, metacognition does have applications in AI and robotics and, by extension, may be applied to LAWS; for example, to enable some online (battlefield) learning, so a LAWS can refine its warfighting tactics for greater targeting accuracy but without learning 'wrong lessons'. On the other hand, the above examples appear to go no further than simply having two separate algorithms or components: a 'core' and a 'meta'.[111] Human metacognition of the kind that applies vague legal standards and resolves conflicting human values arguably goes further than this. As Metcalfe and Son argue, there is a distinction between anoetic, noetic and autonoetic metacognition.[112]

- *Anoetic* metacognition involves stimulus-driven judgements of objects and events that are present in time and space, and with no requirement for self-awareness.

---

[105] Ibid., 94.

[106] R. Savitha, S. Suresh and HJ. Kim, 'A Meta-Cognitive Learning Algorithm for an Extreme Learning Machine Classifier' (2014) 6 Cognitive Computation 253.

[107] Ibid., 261-62.

[108] Preeti Bhargava et al., 'The Robot Baby and Massive Metacognition: Future Vision', *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics*, IEEE *Xplore* (2012) 1 <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6400837> accessed 21 May 2018.

[109] Ibid., 1-2.

[110] Ibid.

[111] Indeed, software suppressors such as Arkin's Ethical Governor may be considered to be a rudimentary metacognitive algorithm.

[112] Janet Metcalfe and Lisa K. Son, 'Anoetic, Noetic, and Autonoetic Metacognition' in Michael J. Beran et al. (eds.), *Foundations of Metacognition* (OUP, 2012), 293-98.

- *Noetic* metacognition involves making judgements about internal representations that are no longer present in time and space – again, with no requirement for self-awareness.
- *Autonoetic* metacognition involves making *self-referential* judgements about internal representations and, in addition, having the *self-awareness* that one is intimately involved. This enables a cognisance of one's remembered past and projected future; of what *has* happened and what *will*, or at least *might* happen, upon taking a particular course of action.

3.2.2.2.3 Machine Metacognition Reconsidered

These three types of metacognition are in ascending order of sophistication. Of particular note is the fact that neither anoetic nor noetic metacognition imply self-awareness, whereas autonoetic metacognition does. This is important, as the metacognitive skills outlined above[113] for their relevance in applying legal standards/principles presuppose a degree of introspection, self-awareness and consciousness, for their effective application in a safety-critical context. For example, in understanding the seriousness of civilian harm, commanders and combatants will not only have consulted the Rules of Engagement (which are like LAWS programming), but will also have empathy for fellow humans; recollections of how seriously more senior commanders and their military legal advisers regard civilian harm;[114] an understanding of how the media, and how their own family and peer-groups view the killing of non-combatants,[115] hence how broader/longer-term civilian harm may cast a shadow on the commander/combatant and his unit.[116] These may all drive a more advanced and probing metacognitive effort to ensure that safety-critical attacks will substantively minimise civilian harm, with intentionality and in line with human values.

[113] See (note and bullet-pointed text accompanying) n 67.

[114] Laura A. Dickinson, 'Military Lawyers on the Battlefield: An Empirical Account of International Law Compliance' (2010) 104 American Journal of International Law 1.

[115] Robert Johns and Graeme AM. Davies, 'Civilian Casualties and Public Support for Military Action: Experimental Evidence' (forthcoming) Journal of Conflict Resolution <https://doi.org/10.1177/0022002717729733> accessed 21 May 2018 (presenting consistent empirical evidence of casualty-aversion by the US and British public, and a corresponding loss of support for military forces).

[116] Ibid.

Thus, with the above (negative) assumption on machine self-awareness,[117] it is arguable that a LAWS will not be metacognitive to LOAC standards. Lacking intentionality, its analytical processing is likely to stop at any formalistic solution that maximises military advantage while minimising civilian harm in the narrow circumstances of an attack, and within the limits of machine perception; even if this causes longer-term civilian harm in a way that no reasonable commander would deem to be a reasonable application of LOAC principles.[118] That said, Metcalfe and Son suggest that machines can, at least in principle, become *self-referential* by being programmed to remember their past, to project into their own future and to take account of things they themselves do and see as they move around their environment.[119] This may permit a form of 'automated introspection', from which a LAWS may better assess its current state of knowledge, the implications of given actions, along with any knowledge gaps and ways to address these, before making lethal force decisions.

A counter-argument, however, is that such 'self-referential' capabilities merely involve the encoding and tagging of data in a way that enables the machine to mimic the thought processes of a self-aware human, but without actually being self-aware;[120] thus remaining "a world away from human thinking for IHL purposes".[121] This compelling point is analogous with Searle's 'Chinese Room' critique[122] of the Turing Test,[123] which posited that a person or machine that merely follows explicit instructions does not necessarily 'understand' his/its task.[124] Accordingly, the ease

---

[117] See 3.2.2.1.2.

[118] See 7.2.3 on the importance of human-machine interaction in the context of proportionality assessments.

[119] Metcalfe and Son (n 112), 298.

[120] Ibid. Indeed, as noted in Chapter 2 (note and text accompanying n 160), AI merely mimics human intelligence via knowledge representation and reasoning, in order to act as a rational agent.

[121] Email from Eliav Lieblich to Maziar Homayounnejad (14 August 2016), on file with author.

[122] John Searle, 'Minds, Brains, and Programs' (1980) 3 Behavioral and Brain Sciences 417 (arguing that symbolic representations of the Chinese language, along with a set of instructions on how to manipulate these, would at best enable a non-Chinese speaker to engage in a *syntactic* process. This might give rise to 'correct' answers and perhaps even a 'human-like' conversation between the non-Chinese speaker and a fluent Chinese speaker; yet it would not imply that the former has a true understanding of the *semantic* contents of the language, as formal symbol manipulations by themselves do not have any intentionality).

[123] Alan M. Turing, 'Computing Machinery and Intelligence' (1950) 59 Mind 433 (positing that if a human operator chats simultaneously online with both a computer and another human, but cannot determine which is which, then the computer has human-level AI and, therefore, has human-level understanding).

[124] Searle (n 122).

with which the adversarial examples discussed in Chapter 2 – each being minor perturbations, yet outside the network's training process – were able to spoof the systems, suggests *a priori* that deep neural networks do not 'understand' their image classification task.[125] The same can be said about perverse instantiation of final goals, in which unconstrained variables are set to extreme values.[126] In both instances, machines clearly do not understand that extreme actions are outside the bounds of human notions of reasonableness, unless this becomes an explicit part of their 'instruction book'.[127]

More specifically, the 'encoding and tagging' critique is based on Humphrey's argument that the internalised concept of a 'self' results in an individual who has both a *mind* and a *concept of his own physical body*; such an individual will therefore strive to preserve and protect the physical body, with a consequent evolutionary advantage.[128] Accordingly, a machine may well be programmed to encode and tag data in a way that *displays* autonoetic metacognition; but it does not follow from this that said machine possesses the "deep and meaningful characteristics of what self-reference means to humans and to their survival".[129] Significantly, this negative conclusion finds support in Arkin's Ethical Governor, whose proof-of-concept suggests that robots can be programmed for "self-sacrifice to reveal the presence of a combatant".[130] This is because, unlike humans with their internalised concept of a 'self', robots do not necessarily have a self-preservation instinct, thus they do not need to protect themselves.[131] Lending further support to this reasoning is Vanderelst and Winfield, whose recent experiments with 'ethical robots' demonstrated how quickly and easily these can turn into the polar opposite – exhibiting inverse, thus *unethical* behaviour – merely by reversing the assignment of maxima and minima functions within the

---

[125] Ian Goodfellow, *Presentation at Re-Work Deep Learning Summit* (24 February 2015) <https://www.youtube.com/watch?v=Pq4A2mPCB0Y> accessed 21 May 2018.

[126] See Chapter 2, (notes and text accompanying) nn 227-232 and 244-245, for examples of perverse instantiation.

[127] In which case, the machine still does not 'understand', it merely recognises that the specified extreme actions are prohibited by its programming.

[128] Nicholas Humphrey, *Seeing Red: A Study in Consciousness* (Harvard University Press, 2006).

[129] Metcalfe and Son (n 112), 298. This is arguably the basis for Walzer's familiar quote that "[f]ear and hysteria are always latent in combat, often real, and they press us towards fearful measures and criminal behavior". See Michael Walzer, *Just and Unjust Wars* (5th ed., Basic Books, 2015), 250.

[130] Ronald C. Arkin, *Governing Lethal Behaviour in Autonomous Robotics* (Chapman & Hall/CRC, 2009), 46.

[131] Ibid., 29.

robot's cognitive machinery.[132] While human soldiers are also known to have their 'internal ethical codes' corrupted, this usually occurs gradually over time, and is often preceded with significant mental degradation.[133] Arguably, broader, human-levels of self-awareness and understanding cannot allow such rapidly incongruous behavioural changes to occur at the 'flip of a switch'. Moreover, without these deeply-ingrained human values, it is unlikely that a LAWS will be able to apply LOAC to the standard required of humans.

3.2.2.2.4 Legal Consequences

With the above in mind, Schachter's brief analysis of the "hypothetical metacognitive ATR" (automatic target recognition) focuses on sophisticated goal-directed behaviours, yet all seem to be essentially *technical* processes.[134] Even so, and within this broadly technical framework, the author concludes that no ATR is fully metacognitive:

> Certain ATRs have some aspects of metacognition, but none so far have a comprehensive ability to strategize, plan, monitor, evaluate, repair, and control itself and its performance.[135]

A fortiori, no ATR or AI can be deemed autonoetically metacognitive, thus no LAWS will be able to apply legal standards to the extent required by the laws of war. However, this does not necessarily negate the lawfulness of narrower forms of lethal autonomy. As the International Committee of the Red Cross (ICRC) has noted:

---

[132] Dieter Vanderelst and Alan Winfield, 'The Dark Side of Ethical Robots' (*AIES 2018*, February 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_98.pdf> accessed 21 May 2018.

[133] See Maziar Homayounnejad and Richard E. Overill, 'Preventing Autonomous Weapon Systems from Being Used to Perpetrate Intentional Violations of the Laws of War', *TLI Think! Paper 8/2018* (2018), 12-13 <https://ssrn.com/abstract=3123254> accessed 21 May 2018 (discussing the Haditha massacre of 2005 and the Mahmudiyah gang rape of 2006).

[134] Bruce J. Schachter, *Automatic Target Recognition* (3rd ed., SPIE Press, 2018), 254 (referring to an ATR's ability to understand its capabilities and limitations; the problem it is trying to solve; the availability and quality of input data; and the ability to assign confidence bounds on its conclusions. The ability to monitor its own health, detect component failures and take these offline, while maintaining system performance. The ability to self-regulate and adjust internal parameters if it detects too many false alarms; or to reduce reliance on a particular sensor if the weather is distorting the quality of input data, while increasing reliance on sensors that are less affected by prevailing conditions).

[135] Ibid.

> [M]achines can and do effectively take decisions that have been delegated to them by humans through their computer programming, and without the need to be 'conscious' or to have human-like intelligence.[136]

To be sure, machines can lawfully execute technical processes that commanders and their battle staffs will have anticipated in their legal assessment of a planned attack. It is the human element in such a system that applies the requisite metacognitive thinking, so long as the autonomous 'choices' on the battlefield remain predictable and within lawful boundaries.

3.2.2.3 The Irrelevance of Human Emotion

Proponents of a LAWS ban tend to emphasise the need for real-time human emotion in applying the LOAC.[137] The reason is two-fold: to identify targetable persons by sensing the emotional states of others (mostly in the case of conduct-based targeting);[138] and to exercise compassion in the broader application of the rules.[139] Both of these points are debatable. First, if ever needed for target identification, sensing human emotional states can now be done more accurately via computational processes, which can be integrated into a LAWS.[140] In any event, emotional states are at most a useful cue, which a soldier *may* wish to take into account.[141] They are *legally* irrelevant for target identification,[142] which hinges more on "the objective risk of harm…based upon objective indicators";[143] or, in the case of status-based targeting, membership of a State's armed forces.[144] Moreover, human emotion is *practically* (as

---

[136] ICRC, 'Views of the International Committee of the Red Cross (ICRC) on Autonomous Weapon Systems', *Working Paper Submitted to the CCW Meeting of Experts on LAWS* (11-15 April 2016), 3 <https://www.icrc.org/en/download/file/21606/ccw-autonomous-weapons-icrc-april-2016.pdf> accessed 21 May 2018.

[137] Human Rights Watch, *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch, 2012).

[138] Ibid., 29 and 31 (assuming only humans can empathise with other humans, in order to detect emotional states).

[139] Ibid., 38

[140] See 2.5.1-2.5.2 on AI and deep learning, especially (note and text accompanying) nn 204-205 therein.

[141] It is acknowledged that in a counterinsurgency setting, emotional cues often take on greater significance, and may be specifically included in the Rules of Engagement. Even so, these are operational, not legal requirements.

[142] Article 51(3), AP I, permits direct attacks against civilians while they "take a direct part in hostilities". This is strictly conduct-based, and imposes no obligation to consider emotional states.

[143] Sassóli (n 47), 333.

[144] Article 43(2), AP I. See also 6.5.2.1.

well as legally) irrelevant to anti-material targeting, to which a LAWS can always be restricted.[145]

Second, the LOAC rules merely have to be applied in accordance with their content and the circumstances at hand.[146] There is no obligation to exercise compassion over and above that which is already contained in the substance of the rules.[147] Besides, the 'compassion' argument overlooks the negative impact of human emotion,[148] such as fear and hysteria,[149] as well as prejudice and the instinct for revenge;[150] all of which can lead to serious civilian harm. Consequently, assumptions about human nature tend to be "mutually offsetting and likely to remain inconclusive" in policy debates.[151] In any event, there is no legal requirement for human soldiers to utilise their emotional senses on the battlefield, nor to require the use of real-time emotional intelligence when appraising the lawfulness of a weapon system.

3.2.2.4 Conclusion on the Absence of Human Qualities

Three broad conclusions can be drawn from the absence of the above human qualities. First, the lack of sentience and self-awareness means that LAWS will not be capable of bearing legal personhood for criminal responsibility, or any kind of individual accountability. As the US Department of Defense (DoD) states in its *Law of War Manual* "[t]he law of war rules…impose obligations on *persons*…not…on the weapons themselves".[152] This, however, is assumed not to pose an A&R gap, so long as accountable humans develop, deploy and operate the machines.

---

[145] As noted in 2.4.4, the 'crawl-walk-run' approach will see LAWS deployments begin with anti-material targeting in uncluttered environments. See also the 'division of labour' argument in 6.5.5.

[146] Sassóli (n 47), 318 ("IHL does not seek to promote 'love', 'mercy' or 'human empathy', but respect based upon objective criteria").

[147] Schmitt (n 58) (noting the importance of observing the military necessity-humanity balance).

[148] See Offices of the Surgeon General, Multinational Force – Iraq and US Army Medical Command, *Mental Health Advisory Team* (*MHAT*) *IV: Operation Iraqi Freedom 05-07: Final Report* (17 November 2006), 34-41 <http://www.combatreform.org/MHAT_IV_Report_17NOV06.pdf> accessed 21 May 2018.

[149] Walzer (n 129), 250 (noting that this often presses soldiers into fearful measures and criminal behaviour).

[150] Homayounnejad and Overill (n 133), 12-13 (discussing the Haditha massacre of 2005 and the Mahmudiyah gang rape of 2006).

[151] Lieblich and Benvenisti (n 31), 256.

[152] US Department of Defense, *Law of War Manual* (DoD, 2015; December 2016 Update) (hereafter, *US DoD Manual*), §6.5.9.3 (emphasis added).

Second, the absence of autonoetic metacognition means that LAWS will be incapable of interpreting and applying LOAC standards to the extent ordinarily required of human combatants and commanders, and this capability gap will likely persist into the foreseeable future.[153] However, this does not mean that autonomous attack can never be lawful;[154] just that its role will have to be constrained to relatively narrow technical processes, which will approximate to an application of specific LOAC rules, in relatively predictable circumstances. Again, the DoD reflects both parts of this position by stating:

> The law of war does not require weapons to make legal determinations, even if the weapon (e.g. through computers, software and sensors) may be characterized as making *factual determinations*, such as whether to fire the weapon or to select and engage a target.[155]

Namely, those deploying and using LAWS must ensure that an attack is lawful, though they may delegate to a weapon system specific targeting actions, which amount to technical processes that have been anticipated in the legal assessment of a planned attack.[156] Holding system reliability constant, the more sophisticated and 'metacognitive' the control software, the more complex and varied the targeting decisions that may be delegated, in more diverse operational environments. Yet, without a capacity for autonoetic metacognition, LAWS cannot 'apply IHL/LOAC' on behalf of any other person, even if the latter retains principal responsibility.

Finally, the absence of real-time human emotion will have no bearing on the lawfulness of LAWS *per se*, though it may necessitate some operational restrictions.

3.3 Implications of Weapons Autonomy for Legal Analysis

Having sketched a detailed account of the technical aspects of machine autonomy in Chapter 2, as well as what this does *not* imply for LAWS in 3.2, it is now necessary to consider what weapons autonomy *does* imply for the legal analysis of LAWS. The

---

[153] James A Reggia, Derek Monner and Jared Sylvester, 'The Computational Explanatory Gap' (2014) 21 Journal of Consciousness Studies 153, 158 (discussing the persistent "lack of understanding of how high-level cognitive information processing can be mapped onto low-level neural computations").

[154] Cf. Lieblich and Benvenisti (n 31).

[155] *US DoD Manual*, §6.5.9.3 (emphasis added).

[156] In that regard, see 4.2 on human-machine teaming, and 7.3.6.1 on front-loading.

following will be in two sub-parts, focusing on the assignment, timing and character of operational decisions; and the machine-operator relationship.

3.3.1 The Effect of Autonomy on the Assignment, Timing and Character of Operational Decisions

Autonomy in the *narrow* critical functions[157] of a weapons system will have some legally significant effects on warfare. The underlying reason, and the key legal distinction between LAWS and other complex military hardware, is that weapons autonomy *operates on the tactical decision to perform a lethal action*, whereas other complex (non-autonomous) systems have an effect only *after* a decision is made by an accountable WO,[158] or *before* it.[159] Thus, machine autonomy will lead to a) the (re)assignment of operational decisions, b) the earlier timing, and c) changes to the character (namely, the specificity and basis) of those decisions.

To put these in context, recall Boyd's 'observe, orient, decide and act' (OODA) loop, which is the model of a combatant's recurring decision-making cycle.[160] Namely, a soldier or WO on the battlefield first *observes* the target and the environment around him, using all of the human senses. Second, he *orients* himself in terms of interpreting the information gathered. Third, he *decides* how to act by weighing the potential courses of action, based on the knowledge accumulated. Finally, the WO *acts*, or 'executes', the decision made.[161] In short, the OODA model "describes the ongoing mental and physical processes involved in observing one's environment and responding to changes therein in pursuit of some goal".[162] In a manually-operated weapon system, all steps in this loop are undertaken by a single human WO. By contrast, the purpose of machine autonomy is to reassign part or all of the loop to a machine, to realise some operational advantage.[163] Indeed, it only takes a moment's reflection to see the OODA loop as an expression of the familiar 'sense-think-act'

---

[157] Recall that 'narrow' means target *recognition* by the weapon system's sensory hardware and control software, rather than the broader targeting *process* that will be discussed in Chapter 5.

[158] McFarland (n 14), 21. For example, precision-guided missiles utilise guidance technologies only after the WO selects a specific target and activates a 'fire' command.

[159] For example, decision-support systems may decide which potential targets to present to the WO, who then selects between them. See 5.5.3 on *Project Maven*.

[160] This was briefly introduced in 2.2.2.

[161] William C. Marra and Sonia K. McNeil, 'Understanding 'The Loop': Regulating the Next Generation of War Machines' (2013) 36 Harvard Journal of Law & Public Policy 1139.

[162] McFarland (n 14), 13.

[163] Ibid. See 1.2.2 on the advantages of machine autonomy.

paradigm of robotics. This now raises the question as to how the substitution of robot for human (or of 'sense-think-act' for OODA) will affect the assignment, timing and character of operational decisions.

3.3.1.1 Autonomy Reassigns Operational Decisions

Recall that machine autonomy involves software-based controllers 'stepping into the shoes' of the human soldier on the battlefield. This effectively transfers part of the burden of the OODA decision-making cycle to the machine, thereby reducing human input in the form of lowered physical and mental interaction.[164] This relieves the human WO, who would otherwise have had to make sense of battlefield intelligence, decide what course of action to take, and to take that action in a timely manner.[165] Yet, this does not mean that no human is making any legally relevant decisions and, given the above technical account of the stored program concept, nor does it mean that the LAWS itself is taking any legally relevant decisions. Instead, decisions involving the narrow critical functions of the weapons system are embodied in software written by developers and programmers, who exercise a form of 'pre-bound discretion'. Legally relevant decisions are also taken by commanders and their battle staffs who deploy the LAWS, and by WOs who activate and potentially monitor its operation.[166] As will be argued in subsequent chapters, such pre-bound discretion is capable of affording the necessary protections demanded by LOAC, if informed by effective programming and a highly deliberative targeting process that incorporates stronger, additional and earlier precautions, before and during deployment.[167]

To summarise: autonomy reassigns the battlefield OODA loop from soldier/WO to LAWS; but since LAWS are incapable of making legally recognised 'decisions', such operational decisions concerning the critical functions of a weapon system are, in effect, made by non-traditional decision-makers. These include programmers, commanders and the battle staffs.

3.3.1.2 Autonomy Necessitates Earlier Timing of Operational Decisions

Reassigning operational decisions from those on the battlefield to those who define the behaviour of the LAWS and deploy it will necessarily mean that key targeting

[164] Ibid., 22.

[165] Ibid.

[166] Ibid.

[167] See Chapters 5-7.

decisions are made *earlier* in the targeting cycle and at locations further away from the 'hot battlezone'.[168] Specifically, these decisions (on *whether* and *how* to perform the narrow critical functions) will have to be made at the time the relevant behaviour is programmed into the machine *and* at the time the decision is made to deploy the LAWS;[169] not necessarily at the time that a concrete situation calling for lethal attack arises.[170] That is, decisions on the use of lethal force can only be made while human programmers, commanders and WOs have the opportunity to adjust system parameters. This is important as it assumes concrete situations arising in armed conflict will not substantially differ from those envisioned at the time the machine was developed, tested and deployed.[171] However, where situations do run the risk of substantially differing, suitable restrictions and precautionary measures should be put in place to prevent unintended engagements on the battlefield.[172]

3.3.1.3 Autonomy Changes the Underlying Character of Operational Decisions

The reassignment of operational decisions will also change the character of those decisions by way of their *generality* and *basis*. On 'generality', human-made decisions that are normally carried out in real-time will be replaced with (or supplemented by) more general programmatic instructions that are fed into the machine's software in advance.[173] This effectively means that individual decisions on the use of lethal force are substituted by broader policy-like choices, which are applicable to the range of situations matching the pre-programmed parameters.[174]

Second, in terms of 'basis', lethal attack 'decisions' taken via a LAWS cannot be based on the observation of concrete situations arising on the battlefield.[175] Instead, they will have to be based on the experience and foresight available at the time the machine is programmed, which is then supplemented with the knowledge and intelligence of the

---

[168] Jeffrey S. Thurnher, 'Means and Methods of the Future: Autonomous Systems' in Paul AL. Ducheine, Michael N. Schmitt and Frans PB. Osinga (eds.), *Targeting: The Challenges of Modern Warfare* (TMC Asser Press, 2016), 178, 193 and 194.

[169] McFarland (n 14), 23. But see 7.2.3.4 on real-time human involvement during battle.

[170] The exception being when there is a dial-in capability for real-time human control. See 7.2.3.4 for a hypothetical example.

[171] McFarland (n 14), 23.

[172] See 6.5.5, 7.2.3 and 7.3.

[173] McFarland (n 14), 23.

[174] Ibid.

[175] Of course, lethal attacks will be *triggered* by visual, radar, sonar and other data-based observations, which meet the parameters for kinetic force; but these will be machine-triggered responses, rather than decisions based on contemporaneous human judgement.

commander, as well as the technical expertise of his battle staffs, when deciding to deploy the weapon.[176]

3.3.1.4 Consequence: Autonomy Weakens the Causal Nexus Between Human Decisions, and Specific Actions and Outcomes

One important consequence of the aforementioned changes is that the causal link between a specific human decision and a specific action or outcome (such as a particular target being engaged or the extent of collateral damage that results) will very likely be weakened when that decision is implemented via an autonomous platform. To reiterate, LAWS will entail decisions on lethal attack being taken by non-traditional decision-makers; in advance of an actual conflict scenario; on the basis of *ex ante* knowledge, intelligence, experience, expertise and foresight; and framed in relatively broad and general terms. Arguably, this leaves much potential for a break in the chain of causation, hence a degree of unforeseeability in the behaviour of a LAWS.

This weakening of the causal link raises the possibility that weapons autonomy will – if not appropriately developed, tested and deployed – lead to unintended engagements and excessive collateral damage on the battlefield. As noted above, the problem will find its solution in the various forms of human control that narrow-loop autonomy will demand,[177] along with a broader notion of precautions in attack[178] incorporated in the broader targeting process.[179]

3.3.2 The Machine-Operator Relationship: 'Delegation', Not Abdication

As detailed above, autonomy on the battlefield requires the human decision-making cycle (the OODA loop) to give way, at least partially, to the 'sense-think-act' paradigm of robotics. A highly autonomous weapon system will execute relatively more of the OODA loop, using only the WO's high-level instruction as guidance in the 'decision' stage of a loop;[180] the result is that the 'nearest' human assumes a 'commander-like' role.[181] Accordingly, autonomy can be seen as shaping the relationship between the WO and the machine.

---

[176] McFarland (n 14), 23. Again, there are exceptions, such as that described in 7.2.3.4.

[177] See Chapter 4, especially 4.4 and 4.5.

[178] See Chapter 7, especially 7.3.5 and 7.3.6.

[179] See Chapter 5, especially 5.3.

[180] McFarland (n 14), 13.

[181] See Figure 2.3 in Chapter 2.

However, this does not imply any formal command responsibility, which itself is partly based on the capacity of another legal entity to commit a violation of LOAC.[182] As explained above, LAWS will be mere tools and no different in law to any other weapon system. Concretely, the WO does not – either in fact or in law – abdicate his LOAC obligations by reason of using a LAWS in combat; rather, part of his combat role is performed by the machine, which executes the WO's high-level instructions. This follows from the above account of LAWS as deterministic machines, which are not truly autonomous entities in a human sense.

As a corollary, LAWS are assumed not to be completely 'independent' machines that operate without any human control: the human-machine relationship is "not severed, it is only modified".[183] Choices made by developers and programmers; by commanders and their battle staffs; and by front-line operators, will all impose constraints on each mission.[184] Accordingly, and despite a lack of narrow-loop supervision, human guidance to a lesser or greater extent will exist to ensure no abdication of control by responsible human beings.

3.4 Conclusion

While Chapter 2 demonstrated the sophisticated, yet technical and brittle nature of LAWS, this chapter has argued that such characteristics will have important legal consequences and non-consequences. Specifically, there are three distinct and broad implications of weapons autonomy during the execution of an attack.

- Lack of agency and legal personhood, for establishing A&R.
- Lack of autonoetic metacognition, for applying broad legal standards and principles.
- Legal assessments and potentially accountable human decisions on lethal force having to be taken earlier, on a more generalised basis, and by non-traditional decision-makers.

---

[182] Chantal Meloni, 'Command Responsibility: Mode of Liability for the Crimes of Subordinates or Separate Offence of the Superior?' (2007) 5 Journal of International Criminal Justice 619; Cf. Peter Margulies, 'Making Autonomous Targeting Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts' in Jens David Ohlin (ed.), *Research Handbook on Remote Warfare* (Edward Elgar, 2017).

[183] McFarland (n 14), 14.

[184] Ibid.

The combined result will be a potential weakening of the causal nexus between human decisions and specific LAWS actions and battlefield outcomes. Accordingly, retaining A&R and ensuring the lawfulness of the use of lethal force will depend on the *type* and *degree* of human control over the actions of a LAWS. With this in mind, Chapter 4 will now examine the 'meaningful human control' concept, as a precursor for the more detailed application of the targeting process and the LOAC rules, which will follow in subsequent chapters.

Chapter 4
‘Meaningful Human Control’ in Autonomy

4.1 Introduction

As the two previous chapters have argued, lethal autonomous weapon systems (LAWS) will neither assume legal obligations/bear legal responsibility, nor have the capacity to draw legal judgments on behalf of their users, in relation to international humanitarian law (IHL)/law of armed conflict (LOAC) targeting rules. Moreover, near-term systems will often operate at the margins of performance: either with super-human accuracy and precision, or with brittleness and potential failure (depending on context and circumstances). These inevitably lead to the question of human control in autonomy; specifically, the *type* and *degree* of such control that will be required in relation to LAWS.

Currently, States parties in Geneva have expressed near-unanimous support for the idea that there must be human control over LAWS, but there is no consensus on what exactly this should entail. Yet, such considerations are important in light of the reassignment, earlier timing and more general character of operational decisions, as these changes risk weakening the causal nexus between human decisions and specific battlefield outcomes. Accordingly, notions of human control have moved centre stage, to ensure militaries and their personnel remain accountable and responsible over the use of force. Whether ‘meaningful human control’ should become a legal standard in and of itself is, however, a different matter.

The following chapter contains four main sections. First, 4.2 explains the practical need for human judgment and control in autonomy, focusing on the different cognitive attributes of humans and machines. Second, 4.3 outlines the structural legal argument for human judgment and control over each ‘individual attack’, and it further argues that this obviates any need to codify a distinct legal requirement for human control; in fact, such a move would distort the crucial military necessity-humanity balance. Third, 4.4 identifies the human-machine interaction ‘touchpoints’ that have emerged from the current Geneva process, and it considers some of their practical and legal implications. Finally, 4.5 distils the substantive elements of human control from several existing contributions. These broadly comprise *predictability and reliability*; setting

operational constraints; and stipulating the *conditions of judgment*. Again, it will be argued here that human control – while undoubtedly useful for the application of the LOAC rules – should not be regarded as a distinct concept, lest this blurs the clarity of existing legal obligations.

While there are many variant terms for human control,[1] the one that has garnered most attention – both in the academic literature and at the current Geneva process – is the Meaningful Human Control (MHC) concept.[2] For consistency, the following will mostly adopt this term.

4.2 The Practical Need for Human Judgment and Control

Humans and computers have different cognitive strengths and weaknesses, and this is brought into sharp focus when we consider the distinction between **automatic** and **controlled** processing.[3] The former refers to the *fast processing* of routine data for deductive reasoning; computers generally perform this better than humans.[4] The latter refers to *slower deliberative processing* for inductive reasoning, recognising novel patterns, autonoetic metacognition and meaningful judgment; as was apparent from 3.2.2.2, humans perform this better than machines.[5] Sharkey argues that only when these attributes are in optimal balance with 'human-machine collaboration' can the weapon system have superior humanitarian impact.[6] Writing with a slightly different focus, Scharre draws a similar conclusion, advocating the idea of a 'centaur warfighter' that will "leverage the precision and reliability of automation without sacrificing the robustness and flexibility of human intelligence".[7] Similar findings are also apparent in other works, specifically on warfare.[8]

---

[1] For example, 'appropriate levels of human judgment' (US), 'intelligent partnership' (UK), and 'minimum standards for meaningful control (*International Committee for Robot Arms Control*).

[2] See Article 36, *Killing by Machine: Key Issues for Understanding Meaningful Human Control* (6 April 2015) <http://www.article36.org/wp-content/uploads/2013/06/KILLING_BY_MACHINE_6.4.15.pdf> accessed 10 May 2018.

[3] Noel Sharkey, 'Staying in the Loop: Human Supervisory Control of Weapons' in Nehal Bhuta et al. (eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (CUP, 2016), 30-34.

[4] Ibid. See also Daniel Kahneman, *Thinking, Fast and Slow* (Penguin, 2012).

[5] Ibid. Recall from 2.2.3.2 that machines need assigned tasks to be both precise and tangible.

[6] Sharkey, ibid. Although the author does not comment on military utility, it is assumed that in many scenarios the humanitarian benefit is coterminous with the military advantage.

[7] Paul Scharre, 'Centaur Warfighting: The False Choice of Humans Vs. Automation', (2016) 30 Temple International & Comparative Law Journal 151, 152.

[8] ML. Cummings, 'Artificial Intelligence and the Future of Warfare', *Chatham House Research Paper* (January 2017), 4-8 <https://www.chathamhouse.org/sites/files/chathamhouse/publications/research/2017-01-26-artificial-

4.2.1 Human Control Over the Machine

Within this framework and in a military weapons context, the human performs three roles.[9]

(1) **Essential operator**, without which the weapon system cannot accurately and effectively complete engagements.

(2) **Fail-safe**, ready to intervene and alter or halt the weapon system's operation, if it begins to fail or if circumstances change such that the engagement is no longer appropriate.

(3) **Moral agent**, making value judgments on the appropriateness of the use of force in a particular attack.

Only the first of these roles can be reliably delegated to a technical process, while the latter two are assumed to be uniquely human domains.[10] That said, in limited circumstances, the 'fail-safe' function can arguably be automated via rudimentary metacognitive algorithms, which can override the main control system;[11] this is the intention behind software suppressors, such as Arkin's Ethical Governor.[12] By contrast, the 'moral agent' role will require more sophisticated traits like autonoetic metacognition and the ability to identify with human values, especially to avoid a perverse instantiation of final goals.[13]

Consequently, while the mechanics of force execution will gradually give way to algorithmic decision-making, two practical realities will need to hold true. First, the wider (deliberate) targeting process will need to embed sufficient checks and balances in which deliberative and professional human judgment shapes all LAWS deployments.[14] Second, for the narrow (dynamic) targeting process, user interface designs and data connectivity should be optimised to enable controlled processing to

---

intelligence-future-warfare-cummings-final.pdf> accessed 9 May 2018 (distinguishing skills, rules, knowledge and expertise along a cognitive continuum in which the first two are easily automated, while the latter two demand human judgment to resolve uncertainty and ambiguity).

[9] Scharre (n 7), 154.

[10] Ibid., 156. Indeed, as demonstrated in Chapter 2, the inherently brittle nature of narrow AI will mean human judgment and control are needed whenever the context for action risks moving outside system parameters, or whenever value judgments are likely to be needed.

[11] See 3.2.2.2.2-3.2.2.2.3.

[12] Ronald C. Arkin, *Governing Lethal Behaviour in Autonomous Robotics* (Chapman & Hall/CRC, 2009).

[13] See 2.5.3.1 and 2.5.3.3.

[14] See Chapter 5, especially 5.3.

pervade the application of lethal force.[15] This is especially important in complex and unpredictable battlefields, which often play to the weaknesses of automatic reasoning;[16] though it may not be possible when actions are required faster than humans can react, or when operating in a communications-denied environment.[17]

4.2.2 Machine Assistance to Enhance Human Control

In delegating more of the 'essential operator' role to a machine, there is a distinction between *control* and *direct manipulation*.[18] While the two may overlap with simple weapons like standard issue rifles or hand grenades, they tend to diverge with more complex systems, and those operating in relatively unpredictable environments. For example, semi-autonomous weapons like precision-guided munitions utilise automatic processing capabilities as an 'intervening mechanism' to counteract adverse environmental conditions, which are objectively programmable, yet move too fast for (human) controlled processing.[19] This ensures the human-selected target is accurately engaged, despite strong winds or evasive manoeuvres by the target. In this sense, human control may be *enhanced* by an automatic mechanism that *decreases* the operator's physical manipulation of the weapon's aiming system.[20] Accordingly, direct manipulation is only contingently related to the degree of control, and much will depend on the nature and extent of external/intervening factors.[21] Should these be too numerous and too fast, effective human control is likely to be enhanced by delegating more physical/execution processes to the machine, while freeing the human operator to exercise more deliberative reasoning and controlled processing in relation to targeting.[22] This of course is the essence of the 'centaur warfighter' approach, which seeks to harness the respective human-machine cognitive strengths.

---

[15] Sharkey (n 3), 34.

[16] Ibid., 32.

[17] Scharre (n 7), 159 (giving the example of defending against saturation attack from missiles and rockets, which are likely to overwhelm human operators; and offensive attacks using swarming munitions).

[18] International Panel on the Regulation of Autonomous Weapons (iPRAW), 'Focus on the Human-Machine Relations in LAWS', *"Focus on" Report No. 3* (March 2018), 13.

[19] See also Chapter 2 (n 45), explaining a similar function on the *Phalanx CIWS* and the *Patriot* missile battery.

[20] iPRAW (n 18), 13.

[21] Ibid. For example, the time-lag between trigger-pull and kinetic effects; how unpredictable weather conditions are in the meantime; whether the target itself will move before kinetic effects, and how fast or unpredictable such movements are likely to be.

[22] This same principle applies in civilian fields, with fly-by-wire and electronic stability programs enhancing the control of aircraft pilots and car drivers, respectively, despite reducing their direct manipulation of the vehicle.

4.2.3 Machine Restrictions on Human Action

Yet, the hybrid focus of the 'centaur warfighter' model should counsel against any misplaced faith in the human, even in relation to controlled processing. While the MHC concept resists the myth of technological infallibility, it risks entrenching a myth of human infallibility instead.[23] Thus, to more effectively combine the cognitive strengths, we should keep in mind that both humans and machines are flawed, albeit in different ways.[24] Accordingly, it is not just human control over a LAWS and the latter's assistance to the operator, but also appropriate *technical restrictions* on human action that will give fuller effect to lawful human intentions.[25] Hence, 'fail-safe' mechanisms like the Ethical Governor are aimed not just at preventing inadvertent machine action, but also unsafe or erroneous deployments by human commanders.[26] Likewise, user interface designs should not just be optimised for controlled processing, but also for automatic vetoes on unsafe or unlawful human action.[27] This has long been the approach in the civilian sector, for example, with Flight Envelope Protection in commercial aviation.[28]

Accordingly, while there is clearly a practical need for human judgment and control in some contexts, there is an equally compelling need for machine action and technical veto in others. As Smith argues, in a well-functioning weapon system "human and machine both empower and limit each other", such that "the fundamental functional question is whether such a system…can remain robust when human or machine elements fail".[29]

---

[23] Bryant Walker Smith, 'Controlling Humans and Machines' (2016) 30 Temple International & Comparative Law Journal 167, 172.

[24] Ibid., 173.

[25] Ibid., 171-73 (giving the examples of sensor-fused weapons, whose submunitions explode only under technically defined conditions).

[26] Arkin (n 12).

[27] iPraw (n 18), 14-15 (discussing 'control by design', which may "create deliberate technical limitations on range or effect as ways to maintain control").

[28] Flight Envelope Protection is a set of technical limits embedded in the control system of a commercial aircraft, which overrides pilot commands that would otherwise force the system beyond its safe operating limits.

[29] Smith (n 23), 176.

4.3 The Structural Legal Argument for Human Judgment and Control Over an 'Individual Attack'

This argument builds on 3.2, in that only human beings can be the addressees of IHL/LOAC.[30] While machines apply a technical process that commanders will have anticipated in their legal assessment of a planned attack, those machines do not have agency (hence, they cannot assume legal obligations); and they lack autonoetic metacognition and broader deliberative thinking (hence, they cannot apply LOAC principles on behalf of persons using/deploying them).

4.3.1 Individual Rules Requiring 'MHC'

By contrast, Article 57(2)(a), AP I, which imposes precautionary obligations on "*those who* plan or decide upon *an attack*",[31] clearly refers to human judgment governing the execution of a single attack. Sub-Paragraph (b) of the same provision stipulates that "*an attack* shall be cancelled or suspended" where it becomes apparent that this has/will become unlawful; the same is written in the restatement of customary law, where the obligation is aptly entitled "Control during the Execution of Attacks".[32] Even before these specific rules are articulated, Article 57(1) sets the tone for an 'individual attack' limitation by requiring that in all military operations, "*constant care* shall be taken to spare the civilian population, civilians and civilian objects".[33] Moreover, Article 52(2) stipulates that a targetable military objective must offer "a definite military advantage" in "the *circumstances ruling at the time*",[34] while Article 51(4)(a)-(b) prohibits as indiscriminate any attack that is not, or cannot be, "*directed* at a *specific* military objective".[35] Sub-Paragraph (c) of that same provision prohibits any weapon "the effects of which *cannot be limited*";[36] namely, those that "have

---

[30] Marco Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified' (2014) 90 International Law Studies 308, 323-24.

[31] Emphasis added. These precautionary obligations are also enshrined in the ICRC's restatement of customary law. See Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law, Vol. 1: Rules* (CUP, 2005) (hereafter, CIHL), Rules 15-18. Rules available at: <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul> accessed 10 June 2018.

[32] CIHL, Rule 19.

[33] Emphasis added. See also CIHL, Rule 15.

[34] Emphasis added. See also CIHL, Rule 8.

[35] Emphasis added. 'Directed' is arguably synonymous with 'controlled', and 'specific' clearly indicates an individual target. The difference between the two Sub-Paragraphs is that (b) focuses on the weapon, while (a) points to how it is used. See also CIHL, Rules 11, 12 and 71.

[36] Emphasis added. See also CIHL, Rules 12 and 71.

uncontrollable effects".[37] Accordingly, human judgment and control over an 'individual attack' are assumed and required by law; in principle and *irrespective of the technical sophistication* of a machine. Notwithstanding, this argument certainly chimes with the current state of the art: as discussed in Chapter 2, machine perception is now relatively advanced in narrow domains, yet it remains brittle in assessing the wider context and in applying 'judgment'. This lends additional (practical) support to the idea that *human* judgment and control must be exercised over individual attacks.

4.3.2 A Broad Standard, Not a Bright-Line Rule

Yet, this legal assumption and requirement does not mean that LOAC imposes specific *ex ante* restrictions on the actions that a LAWS may take. Such restrictions are likely to be contextual, determined (in part) by the technical sophistication of the machine, and the complexity of both its assigned task and the operational environment. Instead, the structural legal requirement merely implies that there are certain boundaries to independent machine operation, based on the notion of an 'individual attack'.[38] That said, an individual attack may potentially comprise multiple acts of violence against multiple specific targets.[39] This is apparent from Article 49, AP I, which defines an 'attack' as "*acts* of violence against the adversary, whether in offence or defence".[40] The use of the plural suggests that an attack can comprise multiple specific engagements,[41] so long as each one is directed at a specific target as per Article 51(4)(a). Indeed, as Anderson explains:

> The size of something that constitutes an attack…doesn't include an entire campaign. It's not a war…[but] an attack is broader than simply the firing of any particular weapon…[it] is going to very often involve many different soldiers, many different units, air and ground forces.[42]

---

[37] Michael N. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) Harvard National Security Journal Features 1, 14 (referring to a biological contagion, which can initially infect a combatant, before spreading to civilians beyond the commander's/operator's control).

[38] Article 36 (n 2); Article 36, 'Key Elements of Meaningful Human Control', *Background Paper to Comments Prepared by Richard Moyes for the CCW Meeting of Experts on LAWS* (11-15 April 2016) <http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf> accessed 10 June 2018. See also 5.5.2 on the risk of this being eroded by developing capabilities.

[39] Ibid.

[40] Also acknowledged in the ICRC's restatement of customary law, CIHL, Rule 1.

[41] Indeed, there are several currently-fielded systems that operate in this way, like the sensor-fused weapon and the multiple-launch rocket system.

[42] Kenneth Anderson, cited in Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (Norton, 2018), 269-70.

Thus, *how many* acts of violence may be considered a single attack is also highly contextual, and will depend partly on the machine's technical sophistication, partly on some of the elements of MHC outlined below, such as the spatio-temporal limits. This is reinforced by the ICRC's *AP I Commentary*, which notes that an attack relates to "a specific military operation limited in *time* and *place*".[43]

4.3.3 The Structural Nature of MHC and the Non-Necessity of Additional Rules

Accordingly, the MHC concept does not add anything substantive to the law, and some argue that it should not be seen as a new concept, much less a distinct legal concept.[44] Conversely, some of the non-governmental organisations (NGOs) cited in 1.1 are explicitly calling for a LAWS ban via codification of the MHC concept, in a Protocol to the Convention on Certain Conventional Weapons (CCW).[45] Several authors have also argued for MHC to be codified as a separate IHL/LOAC principle,[46] both to increase its normative power[47] and to enhance its legal power,[48] via amendments to AP I[49] and the preamble of the CCW.[50] However, as appealing as this may sound, it would disturb the crucial military necessity-humanity (MN-H) balance, and risk

---

[43] Yves Sandoz, Christophe Swinarski and Bruno Zimmermann (eds.), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Convention of 12th August 1949* (Martinus Nijhoff, 1987), ¶ 4783 (emphasis added).

[44] See 'Chair's Summary of the Discussion on Agenda item 6(a) 9 and 10 April 2018, Agenda item 6(b) 11 April 2018 and 12 April 2018, Agenda item 6(c) 12 April 2018, Agenda item 6(d) 13 April 2018', *Chair's Documents at the 2018 GGE Meeting on LAWS* (9-13 April 2018), 5 <https://www.unog.ch/80256EDD006B8954/(httpAssets)/DF486EE2B556C8A6C125827A00488B9E/$file/Summary+of+the+discussions+during+GGE+on+LAWS+April+2018.pdf> accessed 5 July 2018 (summarising State views on MHC at the April 2018 GGE, where some had "pointed out that while terms such as human control did not create an obligation under IHL, their use could be derived from the requirement for compliance with IHL in the application of lethal force"). See also 4.5.6.

[45] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects (adopted 10 October 1980, entered into force 2 December 1983, amended on 21 December 2001) 1342 UNTS 137.

[46] For example, Sharkey (n 3); Elvira Rosert, 'How to Regulate Autonomous Weapons: Steps to Codify Meaningful Human Control as a Principle of International Humanitarian Law', *PRIF Spotlight 6/2017* (November 2017) <https://www.hsfk.de/fileadmin/HSFK/hsfk_publikationen/Spotlight0617.pdf> accessed 10 June 2018.

[47] Rosert, ibid. (for example, by formalising greater considerations of humanity, by raising its prominence via inclusion in key IHL documents, and by allowing it to benefit from the strength of other key IHL principles).

[48] Ibid. (for example, by requiring States to formally consider MHC in their legal reviews of new weapons).

[49] Ibid. (specifically arguing for an amendment to Article 35, to "prohibit…weapons operating without [MHC]"; and to Article 57, to require that "[MHC] is ensured in the conduct of military operations at all steps").

[50] Ibid. (in particular, "Declaring that it is prohibited to employ means of warfare operating without [MHC]").

undermining both administrability and respect for IHL/LOAC by the armed forces that have to implement it.[51] Indeed, issues of humanity (as well as necessity) already undergird all the individual rules, and it is not appropriate to supplement the latter with any further norms that swing the balance in favour of humanity. As Schmitt argues, "the requisite balancing has already taken place",[52] and any further shift in emphasis "risks destabilizing the delicate balance that preserves the viability of IHL in a state-centric normative architecture".[53] Reeves and Thurnher make similar points, specifically in relation to LAWS and the push by NGOs for a pre-emptive ban.[54] In particular, the authors argue that organisations like *Human Rights Watch* are "focused solely on humanitarian considerations", and with limited military expertise, this "hinders their ability to assess properly where the appropriate balance lies".[55] Consequently, NGOs have an "inherent inclination to blindly support humanitarian principles", and "States should not be pressured into foreclosing [LAWS development or deployment] options without conducting a comprehensive review".[56]

Concerns regarding the erosion of the MN-H balance mostly emanate from international courts and tribunals, and from undue pressures exerted by United Nations (UN) bodies, or misinformed but well-mobilised NGOs.[57] Purely State-driven codification towards greater humanitarian concern is, by definition, in keeping with an appropriate MN-H balance, as States are directly affected by both prongs.[58] In a LAWS context, the more likely risk of a disruptive codification is undue pressure from NGOs on States. However even this may be unlikely, as treaty amendments are procedurally

---

[51] Michael N. Schmitt, 'Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance' (2010) 50 Virginia Journal of International Law 795 (explaining that both military necessity and humanity exist in a "fragile equipoise" in IHL, undergirding every individual rule).

[52] Ibid., 839.

[53] Ibid., 796.

[54] Shane R. Reeves and Jeffrey S. Thurnher, 'Are We Reaching a Tipping Point? How Contemporary Challenges are Affecting the Military Necessity-Humanity Balance', *Harvard National Security Journal Features* (24 June 2013), 6-9 <http://harvardnsj.org/wp-content/uploads/2013/06/HNSJ-Necessity-Humanity-Balance_PDF-format1.pdf> accessed 10 June 2018 (arguing that the current LAWS debate "illustrates the aggressive attempts to contravene the principle of military necessity by those who are singularly focused on humanitarian considerations").

[55] Ibid., 8.

[56] Ibid., 8-9.

[57] Schmitt (n 51), 816-29.

[58] Ibid., 838 (arguing that this makes States uniquely placed to perform the balancing task, whereas NGOs focus exclusively on humanity and "they pay no price for forfeiting a degree of military necessity").

difficult to pass,[59] and with the current CCW membership strongly in disarray over how to move forward with LAWS (and some States rather unresponsive to NGO pressures),[60] proposals to codify an MHC principle are arguably unlikely to materialise.

4.4 Human-Machine Interaction 'Touchpoints'

To focus discussions at the 2018 GGE meeting, the Chair synthesised the various State contributions on human-machine interaction 'touchpoints', and derived the following main points:[61]

- Research and development.
- Testing and evaluation (T&E), verification and validation (V&V), and legal reviews of new weapons.
- Deployment, command and control.
- Use and abort.

In each one, it is expected that some deliberative human judgment can be exercised over (actual or potential) weapon systems. After discussions on the floor, States added three more touchpoints: the political level, training, and battle-damage assessment.[62] This derives the following illustrative list, in chronological order.

Touchpoint	Personnel	Illustrative Activities
The Political Level	• Political leaders • Secretaries of Defence • North Atlantic Council	• Deciding on defence procurement priorities • Providing political direction to Joint Force Commanders, to kick-start a deliberate targeting process (see 5.3.1)
Research and Development	• Project managers	• To be fully informed by legal obligations, which should be decisive in programming, weapons design and user

[59] See Article 97, AP I, which requires consultation with all State parties just to decide whether a conference should be convened to consider any amendment.

[60] See, for example, Denise Garcia, 'Governing Lethal Autonomous Weapon Systems', *Ethics & International Affairs* (13 December 2017) <https://www.ethicsandinternationalaffairs.org/2017/governing-lethal-autonomous-weapon-systems/> accessed 10 June 2018 (explaining that State parties are split along three major lines: those wanting a ban or moratoria; those opposing a ban, or even any specific regulation; and those advocating a politically binding agreement based on MHC concepts, but no legal solution).

[61] See 'Chair's Summary' (n 44), 4.

[62] Ibid., 5; author's notes, *April 2018 GGE Meeting on LAWS* (11 April 2018).

	• Engineers from both defence contractors and ministries of defence • Military personnel, especially in T&E	interface designs (e.g. specific instruments and software procedures to enable human input and intervention; recordability, auditability and explainability) • Collaboration between software and hardware designers, to ensure compatibility and reliability • Conduct T&E and V&V procedures, fit for autonomous systems and in realistic settings
The Legal Review of New Weapons, Means and Methods	• Legal review panels: military lawyers, engineers, service personnel, acquisition managers, etc.	• Consider T&E and V&V results • Imposing deployment and use restrictions in line with the testing and technical data • Assessing the system's capacity to be used in accordance with the LOAC
Training	• Human resource personnel • Training personnel (e.g. instructors)	• Training of existing personnel: commanders, legal advisers, battle staffs and weapons operators; to ensure their ability to use a given system in compliance with international law (see 7.3.6.7) • Inclusion of new personnel in battle staffs, e.g. software engineers and roboticists (see 7.3.6.8)
Deployment, Command and Control	• Joint Force Commanders • Unit commanders • Military legal advisers • Various battle staffs: intelligence analysts, weaponeers, etc.	• Deliberate targeting process (see 5.3.1) • Double principle of command and subordination ➢ Framing, re-definition and adjustment of a weapon system's mission to be done by humans only ➢ Concrete decisions on 'when and where' of lethal force to be taken only by humans • Communications links – even if intermittent – must be maintained between the chain of command and the weapon system, to maintain sufficient control and ensure humans take ultimate decisions on the use of force
Use and Abort	• Weapons operators	• Activate the LAWS: know the system's characteristics; ensure these are apt for the operational environment; have sufficient and reliable information on

		both, to make conscious decisions and ensure legal compliance (see 5.3.1.5) • Monitor the LAWS operation with a two-step approach for maintaining control ➢ Understand situation and context, especially when malfunction or battle space changes ➢ Retain the option to intervene/override the system at all stages, or at least during target selection and engagement • All the above implies WO must: ➢ Provide timely data inputs to ensure actions correspond with operator's intention ➢ Fully utilise user interfaces and ensure adequate connectivity during each mission
Battle-Damage Assessment	• Weapons operators • Military legal advisers • Select battle staffs	• Assess for LOAC compliance (see 5.3.1.6 and 7.3.6.6) • Measure any gaps between T&E and field use • Discover possible humanitarian improvements to the LAWS, even if no norms are violated

Table 4.1: Human-machine interaction touchpoints. Source: Author's expanded notes, *April 2018 GGE Meeting*.

4.4.1 Upstream *versus* Downstream Touchpoints

A convenient way to categorise the above stages is to divide them between *upstream* (the first four) and *downstream* (the last three) touchpoints; or *ante bellum* and *in bello*.[63] Upstream touchpoints occur at times and locations far removed from any concrete operation or civilian risk situation (hence, *ante bellum*), whereas the downstream touchpoints all occur during, or at least relatively near to, a battlefield situation (*in bello*). Related to this, the ICRC has noted that it is in the penultimate touchpoint – when a LAWS is in use, and it autonomously executes the narrow critical functions – that "the important question arises as to whether human control in the [preceding] stages is sufficient to overcome minimal or no human control at this last

[63] Heather Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons', *Briefing Paper for Delegates at the CCW Meeting of Experts on LAWS* (11-15 April 2016), 3 <http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf> accessed 7 July 2018.

stage".[64] In this connexion, the views of States and other commentators differ widely on whether a human should authorise specific actions;[65] monitor them with an override and abort capability;[66] rely on upstream programming or downstream (commander) setting of deployment parameters;[67] or insist on a combination of all touchpoints.[68] As will be seen in Chapters 6 and 7, the answer is likely to be context-dependent.

4.4.2 An Individual, Not a Cumulative Standard

Two opposing conclusions can be drawn from Table 4.1. On the one hand, the multiple stages at which human judgment can come to bear may afford a greater *overall* opportunity to ensure (practical) human control, as implied by the ICRC's question above. On the other hand, the involvement of various actors potentially muddies what is (legally) meant by 'MHC over LAWS', raising the temptation to regard 'human control' as being a cumulative standard.[69] This would be problematic, as there can be no 'shared responsibility' over the use of a weapon system,[70] much less can there be shared liability in international criminal law (ICL), which specifically requires that convictions be based on individual criminal responsibility;[71] save in cases of joint criminal enterprise.[72] Moreover, since there is a direct relationship in international law between 'control exercised' and 'responsibility',[73] there can be no 'shared control'

---

[64] ICRC, 'Views of the International Committee of the Red Cross (ICRC) on Autonomous Weapon Systems', *Working Paper Submitted to the CCW Meeting of Experts on LAWS* (11-15 April 2016), 3 <https://www.icrc.org/en/download/file/21606/ccw-autonomous-weapons-icrc-april-2016.pdf> accessed 10 June 2018.

[65] For example, Thompson Chengeta, 'What Level of Human Control Over Autonomous Weapon Systems is Required by International Law?' *EJIL: Talk!* (17 May 2018) <https://www.ejiltalk.org/what-level-of-human-control-over-autonomous-weapon-systems-is-required-by-international-law/> accessed 10 June 2018.

[66] Kjølv Egeland 'Lethal Autonomous Weapon Systems under International Humanitarian Law' (2016) 85 Nordic Journal of international Law 89, 102-03.

[67] For example, Schmitt (n 37).

[68] For example, Article 36 (n 38); Roff and Moyes (n 63).

[69] Thompson Chengeta, 'Defining the Emerging Notion of 'Meaningful Human Control' in Weapon Systems' (2017) 49 New York University Journal of International Law and Politics 833, 865 (referring to the US delegation putting this idea forward at the 2014 CCW Expert Meeting).

[70] Ralph G. Steinhardt, 'Weapons and the Human Rights Responsibilities of Multinational Corporations' in Stuart Casey-Maslen (ed.), *Weapons Under International Human Rights Law* (CUP, 2014), 531-32.

[71] Article 75(4)(b), AP I; CIHL, Rule 102.

[72] Joint criminal enterprise is where a number of persons (e.g. a designer, programmer, commander and a weapons operator) form an organised group to make use of a LAWS in a way that violates ICL. So long as the crimes are committed within the common plan or purpose, each member of the group is individually responsible: *Prosecutor* v. *Tadić* (ICTY Appeals Judgment) IT-94-1-A (15 July 1999), ¶¶ 220, 227 and 228.

[73] Kristen E. Boon, 'Are Control Tests Fit for the Future? The Slippage Problem in Attribution Doctrines' (2014) 15 Melbourne Journal of International Law 329; Amy Tan, 'Responsibility and Control in International Law and Beyond', *The Hague Institute for Global Justice* (27 June 2013)

either. Legally, each actor will have 'control' over a LAWS only in their own capacity.[74]

However, this does not mean there cannot be cumulative or collective control *within* a single touchpoint. For example, in 'deployment, command and control' (the deliberate targeting process) there is often a large number of battle staffs contributing towards 'target development' and 'capabilities analysis'.[75] However, as all their activities occur under the same command and control structure, with an identifiable Joint Force Commander (or unit commanders), it is appropriate to take a relatively more expansive view of MHC here.[76]

Thus, while all actors at all touchpoints do potentially play an important role in ensuring that LAWS are ultimately used lawfully, MHC cannot *legally* constitute the sum total of their activities. Rather, it is a *discrete* standard that applies to each individual entity, albeit collectively *within* each touchpoint entity.

4.4.3 Core and Derived MHC in Law

Nonetheless, in *practical* terms 'control' is more useful as an overall structure than as a standalone concept.[77] This counsels in favour of seeing *complementary* discrete standards that ensure individual accountability while promoting overall human control over LAWS. Arguably, the initial focus (to determine a 'core' MHC standard) should be on the combatant/weapons operator (WO), for three main reasons. First, most writings on the 'accountability gap' focus on the WO's role, arguing that autonomy without his control would make it difficult to ascertain intention for the purpose of establishing criminal responsibility.[78] Thus, beginning with the WO is an important step in addressing this gap, and it chimes with the basic legal inquiry for 'effective

<http://www.thehagueinstituteforglobaljustice.org/latest-insights/latest-insights/news-brief/responsibility-and-control-in-international-law-and-beyond/> accessed 10 June 2018.

[74] Chengeta (n 65); (n 69), 866.

[75] As will be seen in 5.3.1.2 and 5.3.1.3, these two phases involve distributed decision-making in selecting specific targets (or target sets) and matching these to appropriate weapons, such as LAWS.

[76] Merel AC. Ekelhof, 'Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting' (2018) 71 Naval War College Review 61.

[77] Bryant Walker Smith, Lawyers and Engineers Should Speak the Same Robot Language' in Ryan Calo, A. Michael Froomkin and Ian Kerr (eds.), *Robot Law* (Edward Elgar, 2016).

[78] This is true of Human Rights Watch and Wagner, both cited in Chapter 3, n 55 therein.

control' in the various branches of international law, which has been suggested as: "who is the *aggregator of power*, who can be held *accountable*…?"[79]

Second, as the end-user of a LAWS, the WO makes terminal ('trigger-pull') choices that have concrete effects on the battlefield.[80] By contrast, other actors like designers and programmers are far-removed, and their decisions are made in a relatively more abstract setting. Hence, much of the structural legal requirement for MHC over an 'individual attack' directly applies to the WO,[81] again underscoring why MHC standards should begin here.[82]

Finally, once the WO's core MHC standard is established this conveniently determines the responsibilities of all other actors,[83] whose 'derived' MHC standards must support the WO's core obligations. For example, if it is deemed that WOs should conduct periodic battle-damage assessments to keep a long deployment within the individual attack limitation,[84] designers will have to ensure this capability, for example, by designing-in cameras with full-motion video recording and reliable datalinks. In some cases, derived MHC obligations may involve technically 'enforcing' the core standards through further design features. Thus, if it is determined that some forms of online learning[85] would violate Article 51(4)(b), AP I, by introducing the risk of learning 'wrong lessons',[86] programmers may have to hard-code a technical prohibition against online learning combined with target engagement authority.[87] Alternatively, if the capability for online learning is required (for very narrow circumstances), programmers may enable it subject to commander authorisation via strict access control mechanisms.[88] In short, the standard of MHC that WOs are legally

---

[79] Chengeta (n 69), 876; Tan (n 73) (emphasis added). It should be noted, that 'effective control' is referred to in the broadest sense. There is no suggestion that effective control tests developed in specific settings, such as to establish State responsibility or command responsibility, are applicable to LAWS.

[80] Chengeta, ibid., 868. Namely, he is the 'aggregator of power' at the time closest to weapons release.

[81] See 4.3.1. For example, Articles 51(4)(a), 52(2) and 57(2)(b), AP I, all apply just before 'trigger-pull'.

[82] Note that some MHC-related rules, such as those contained in Article 57(2)(a), AP I, directly apply to commanders who also assume 'core' MHC obligations.

[83] Chengeta (n 69), 869.

[84] See 7.3.6.6 on upper engagement limits.

[85] See 2.5.1.4 on potential types/outcomes of online learning.

[86] See 2.5.3.2 on brittleness and the risk of learning 'wrong lessons'.

[87] On technical prohibitions to promote good faith commander intentions, see Arkin (n 12).

[88] Maziar Homayounnejad and Richard E. Overill, 'Preventing Autonomous Weapon Systems from Being Used to Perpetrate Intentional Violations of the Laws of War', *TLI Think! Paper 8/2018* (2018), 38-39 <https://ssrn.com/abstract=3123254> accessed 21 May 2018 (describing various access control

obliged to exercise should serve as a yardstick and guidelines for the technical capabilities and limitations of a LAWS; and for the role of all other actors, which must be congruent with the WO's MHC.[89]

4.5 The Elements of MHC

As noted in 2.3.4.3, a likely form of MHC to moderate weapons autonomy will be through human-prescribed *target parameters*, and *geographical* and *temporal boundaries*, which are set tightly enough to enable control over an 'individual attack'.[90] These 'building blocks' were first popularised by UK-based NGO *Article 36*,[91] which subsequently added the *operational environment*.[92] All else being equal, they restrict autonomous lethal targeting to competences in which machines excel and are more reliable; namely, those involving automatic processing. Importantly, these parameters approximate to some of the criteria and sub-criteria for reliable autonomy introduced in 2.2.3.2, thus they may be expected to make LAWS operations more predictable and controllable for the human commander/WO.

4.5.1 Article 36: MHC Building Blocks in More Detail

Target parameters are (object) characteristics that are amenable to detection by automatic target recognition (ATR).[93] Accurate detection is achieved through a target's infrared, radar or acoustic signatures;[94] and/or its shape and image, among other things.[95] Such detectable characteristics are *proxy indicators* of an intended and legitimate target, which aim to ensure that no objects other than intended targets are selected for lethal attack.[96] Accordingly, the number and type of proxy indicators that are programmed into a LAWS can affect the degree of MHC afforded to a commander/WO. For example, 'motorised vehicle with engine heat signature' may be

mechanisms that ensure high-level authorisation and accountability for sensitive adjustments to military systems).

[89] Chengeta (n 65).

[90] See 4.3.

[91] Article 36, 'Structuring the Debate on Autonomous Weapons Systems', *Memorandum for Delegates to the Convention on Certain Conventional Weapons (CCW)* (14-15 November 2013) <http://www.article36.org/wp-content/uploads/2013/11/Autonomous-weapons-memo-for-CCW.pdf> accessed 10 June 2018.

[92] Article 36 (n 2).

[93] See 2.5.4 on the various approaches to ATR.

[94] This is true of cooperative targets. See 2.5.4.2.

[95] Article 36 (n 2), 4. This is potentially true of both cooperative and non-cooperative targets. See 2.5.4.2.

[96] Article 36 (n 91), 2.

intended to select and engage the military vehicles of the enemy. However, with such a broad target parameter, commanders may deem the risk of inadvertent attacks on civilian and medical vehicles to be unacceptable. By contrast, very specific models of military vehicle that are known to be used only by the enemy – for example, '*T-80* Tank' – combined with appropriate image recognition capabilities,[97] will narrow down the 'tolerance of error'.[98] In turn, this may increase the likelihood that the commander's/WO's intentions are carried out on the battlefield, thereby increasing the degree of MHC over the LAWS.

The **geographical area** over which, and the **time** during which, an attack takes place can also confer MHC by imposing spatial and temporal limitations on the autonomous operation of a LAWS. If the geographical location of the target area is relatively small and fixed in space, and the time window is relatively brief, a human commander/WO will be more likely to possess the necessary information, at any given moment, to determine which objects other than legitimate and intended targets may be at risk of being targeted by a LAWS, or incidentally affected by an attack.[99] Moreover, narrow spatio-temporal boundaries reduce the likelihood that the physical environment will change during a deployment, and degrade sensory perception.[100]

The **operational environment** in which a LAWS is deployed can also confer MHC by affecting the 'tolerance of error'.[101] Thus, in a cluttered environment where there is a dense concentration of civilians and civilian objects that potentially match the target parameters, and which are sufficiently nearby as to be affected by an attack, commanders are likely to have less control over the unintended consequences of a LAWS; hence MHC may be undermined. By contrast, in a traditional battlespace – air, sea or open desert – civilians are less likely to be present, hence commanders are more likely to exercise MHC in the deployment/use of LAWS.[102]

[97] See 2.5.2 on object recognition and 2.5.4.1 on standard ATR approaches.
[98] Article 36 (n 2), 4.
[99] Article 36 (n 91), 3.
[100] See 2.5.5.1.
[101] Article 36 (n 2), 4.
[102] Ibid.

To summarise, the potential lawfulness of a LAWS deployment hinges on *limitations* being placed on its independent operation. These enable those responsible for the planning and conduct of an attack to make informed judgments on its military utility and necessity, and on broader issues regarding the legality of the use of force.

4.5.2 Article 36: Key Elements of MHC

More recently, *Article 36* discussed a more comprehensive and consolidated understanding of the 'key elements' of MHC.[103] These are necessary to allow for the effective application of the LOAC rules and to prevent the structure of the law from progressive erosion.[104] They comprise the following.[105]

(1) ***Predictable, reliable*** **and** ***transparent*** **technology**. Together, these imply that the technology will follow a discernibly repeatable pattern; will not be prone to failure, and will be designed to fail safe; and will be readily understandable to those deploying and using it. These characteristics have implications for (upstream) design and testing, as well as the legal review of new weapons, and their intended result is to increase the level of downstream control exercised by commanders/WOs.

(2) **Accurate information for (and understanding by) the user on the *outcome* sought, the *technology* and the *context* of use**. Focusing on the downstream, the crucial issue of predictability is determined not just by the technology itself, but also by the commander's *understanding of the technology*[106] and how it will interact with the operational environment. Thus, information on *context* is also needed, especially on the presence and movements of civilians and civilian objects. Furthermore, the ability to understand context is directly linked to both the size of the area in which a LAWS will operate, and the time during which it will operate: for any given environment, a greater area and longer duration of autonomous operation reduces predictability and human control. Yet, different operational environments – land, air or sea; sparse or cluttered – have different characteristics: a large geographical area on the high-seas may afford

[103] Article 36 (n 38), 4.

[104] Ibid., 3. This refers to potential pressures arising from autonomous technologies to expand the concept of an 'attack', and will be explained further in 5.5.2.

[105] All summarised from ibid., 4.

[106] For example, its target profile templates, the nature and extent of its multisensory phenomenologies (on which, see 2.5.4.1), how it will apply kinetic force, etc.

greater contextual understanding than a smaller area on land.[107] Consequently, an understanding of both technology and context should enable the commander to assess likely *outcomes* from an attack, including both the objectives sought and the unintended (collateral) effects. This enables the commander and his legal advisers to a) assess the validity of a military objective at the time of attack, and b) to evaluate and select a proposed attack option within the LOAC rules. The final assessed outcome must match the commander's lawful intent with a certain degree of probability. If it does not, then the time and space of operation should be reduced until an acceptable degree of certainty can be discerned. Ultimately, these all enhance the predictability of outcomes – both positive and negative – that form the basis of legal assessments.

(3) ***Timely* human *judgment* and action, and a potential for *timely intervention***. Based on the information and understanding gleaned from the second key element, human commanders need to apply their judgment – as implied by the structural legal analysis in 4.3.1 – and select a specific weapon system, to be deployed in a specific way. Timeliness is of vital importance here, as the information being acted upon becomes less relevant over time. Also, given the often-fast tempo of warfare, timely intervention may be needed to minimise the risk of unintended engagements.

While these key elements represent further development of the four parameters of MHC articulated by *Article 36* in earlier years, they do not provide exact boundaries or any bright-line rules. Instead, the MHC elements should be seen as providing a framework within which the application of the LOAC rules should be articulated within the targeting process.[108]

4.5.3 Horowitz and Scharre: Essential Components of MHC

A similar approach to the above can be seen in Horowitz and Scharre, whose *Primer* on MHC discussed three 'essential components' of the concept, which are derived from the kind of human control exercised over present-day weapon systems. The authors argue this is necessary to ensure MHC reflects the realities of the battlefield

[107] Note that this coincides with simpler operational environments being more amenable to autonomous action, as explained in 2.2.3.2.

[108] See Chapter 5.

and how weapons are actually used, thereby increasing the likelihood that it will be taken up by key stakeholders.[109] The essential components are:

(1) Human operators are making *informed, conscious* decisions about the use of weapons.

(2) Human operators have *sufficient information* to ensure the lawfulness of the action they are taking, *given what they know* about the *target*, the *weapon* and the *context* for action.

(3) The weapon is *designed and tested*, and human operators are *properly trained*, to ensure *effective control* over the use of the weapon.[110]

The intention is for MHC to ensure that commanders are making conscious decisions, with sufficient information to *remain legally accountable* for their actions.[111]

4.5.4 ICRC: Distilling MHC From Current-Day Weapon systems

The ICRC has also extracted a number of MHC factors from existing weapon systems that select and engage targets without human intervention.[112] It concludes that human control in these systems is determined by:[113]

(1) **Verified technical performance** of the weapon system for its intended use, as determined during the development stage.

(2) **Sufficient knowledge and understanding of the weapon system's functioning** by the commander/WO, and **adequate situational awareness** of the operational environment, especially in relation to civilian risk. This will form the basis of **operational parameters**, which are largely designed-in at the development stage, and further set or adjusted at the activation/deployment stage. These include constraints on:

- The *task assigned* to the weapon system.
- The *type of target* it may attack.
- The *type of force* and *munitions employed* (and their associated effects).

---

[109] Michael C. Horowitz and Paul Scharre 'Meaningful Human Control in Weapon Systems: A Primer', *CNAS Project on Ethical Autonomy Working Paper* (March 2015), 10 (emphasis added) <https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf> accessed 10 June 2018.

[110] Ibid., 14-15.

[111] Ibid., 15.

[112] ICRC (n 64).

[113] The following is summarised from Neil Davison, 'A Legal Perspective: Autonomous Weapon Systems Under International Humanitarian Law' in UNODA, *Occasional Papers No. 30: Perspectives on Lethal Autonomous Weapon Systems* (United Nations, November 2017), 11-15.

- The *operational environment* in which the weapon system is to operate.
- The *mobility* of the weapon system in space.
- The *time frame* of its operation.
- The *extent of adaptability* permitted.

(3) The **level of human supervision** and **ability to intervene** during the operation stage. This supplements the parameters set on deployment, to continue retaining MHC.

Much of this overlaps with 4.5.1 and 4.5.2, though four (new) points need elaborating. First, constraints on the *task assigned* can be quantitative (limiting these to one or a few tasks), or qualitative (limiting these to the most precise, well-defined tasks with tangible outcomes).[114] Second, constraints on the *type of target* refers to whether persons, vehicles and/or broader categories of objects are targeted.[115] Third, the *type of force* and *munitions employed* refer to whether lethal or non-lethal force is used; and if the former, how large the blast radius and potential collateral effects. Finally, *adaptability* can be of two kinds: a) setting own goals, or adapting human-set goals in response to the environment;[116] or b) being deployed in online learning mode to refine existing tactics, or to develop new targets and tactics on-the-fly.[117]

4.5.5 Common Strands and Elements of MHC

Others have also enumerated MHC elements, much of which overlap with the above.[118] Overall, the concept arguably comprises three distinct but complementary sets of elements:

- **Predictability and reliability** of systems and their use, by way of weapons design, V&V, T&E, learning parameters, legal review and personnel training.[119]

---

[114] Ibid., 2. See 2.2.3.2. An example of a precise task may be defending against incoming projectiles of a defined speed, distance and trajectory. The tangible outcome would be the projectile being destroyed or averted.

[115] Ibid.

[116] For example, where an anti-material system comes under attack by a person, it may be authorised to assess the situation and determine whether returning fire (thus, shifting to anti-personnel targeting) is warranted.

[117] ICRC (n 64), 3.

[118] For example, iPRAW (n 18), 14-15; and UNIDIR, 'The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward', *UNIDIR Resources, No. 2* (2014), 5-7 <http://www.unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf> accessed 10 June 2018.

[119] Weapons law is outside the scope of this thesis; however, see 6.3 and n 30 therein.

- **Operational constraints** on time and space,[120] targets,[121] nature and quantity of tasks,[122] the type of force and munitions employed,[123] the operational environment,[124] and the extent of adaptability;[125] all to keep within the 'individual attack' limitation.
- **Conditions of judgment**, in particular, requiring informed and conscious decision-making, based on sufficient information about the applicable law, the target, the weapon, the operational environment, the context for action and the interaction between these.[126] This includes incorporating a feedback loop that enables a commander/WO to detect when deployment and use are straying from lawful boundaries, and to abort the mission.[127]

The last two bullet points concern downstream MHC, which applies during the targeting process. The first bullet point is concerned with upstream MHC, which generally occurs far from any contact zone, but is vital for enabling MHC during combat. Importantly, there is no 'one size fits all' formulation. The extent of human control required in any given mission must be adapted to the specific task and environment at hand, and it must allow the human to make meaningful decisions that comply with the LOAC rules.[128]

4.5.6 An Interpretive Aid, Not a Distinct Legal Concept

Despite its importance for legal compliance, MHC should arguably not be seen as a distinct concept. In fact, it may be harmful if it is treated discretely, lest it becomes the primary object of a weapons deployment and potentially misapplied in a way that conflicts with LOAC norms. To be sure, the concept is useful in that it brings to the fore issues that are crucial to the lawful use of LAWS, but which have been unnecessary to consider with manned and remotely-piloted systems. Namely, it provides a suitable lens through which to evaluate how a given deployment will meet LOAC norms. Moreover, MHC is "intuitively appealing" to all States in the Geneva

[120] See 7.3.6.4 on spatio-temporal limits.
[121] See 7.3.6.5 on target parameters.
[122] See 7.3.6.6 on upper engagement limits.
[123] See 6.5.5 on Canning's proposal.
[124] See 6.5.5 and 7.2.3.1 on restricting deployments to benign operational environments.
[125] See 7.3.6.3 on a precautionary approach to online learning.
[126] See 7.3.6.7 on training, and 7.3.6.8 on staffing.
[127] See 7.2.3.4 on real-time communication links.
[128] 'Chair's Summary' (n 44), 6.

process,[129] as it affords a degree of latitude that accommodates different views while identifying bipolar extremes for consensus.[130] Importantly for LAWS, the flexibility of the term allows for adaptive interpretations over time, while ruling out extreme and dangerous developments, and this makes it particularly useful for addressing new technology.[131]

However, there are concerns that if MHC was to evolve into a new legal standard in and of itself, it may risk blurring the clarity of existing laws, and this scepticism is shared by both academics[132] and States[133] alike. The argument is similar but distinct to the one raised in 4.3.3, which focused on MN-H distortions in *future*. The current argument concerns how a distinct MHC concept may cause *uncertainty* of *existing* legal standards, to the extent that civilian risk and force casualties may *increase*. For example, consider the 'minimum necessary standards for meaningful control' put forward by the *International Committee for Robot Arms Control* (*ICRAC*):

> First, a human commander (or operator) must have full contextual and situational awareness of the target area and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack.
>
> Second, there must be active cognitive participation in the attack and sufficient time for deliberation on the nature of the target, its significance in terms of the necessity and appropriateness of attack, and likely incidental and possible accidental effects of the attack.

---

[129] UNIDIR (n 118), 2.

[130] Rebecca Crootof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 Temple International & Comparative Law Journal 53, 56 (noting that this can be beneficial for securing early State support, so long as consensus on the details does later materialise).

[131] Colin B. Picker, 'A View from 40,000 Feet: International Law and the Invisible Hand of Technology' (2001) 23 Cardozo Law Review 149, 184-87 (discussing the glacial pace of law-making in technology-related fields).

[132] For example, Horowitz and Scharre (n 109); Crootof (n 130); Thilo Marauhn, 'Meaningful Human Control – and the Politics of International Law' in von Heinegg, Frau and Singer (eds.) (n 21); Kenneth Anderson and Matthew C. Waxman, 'Debating Autonomous Weapon Systems, their Ethics, and their Regulation under International Law' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds.), *The Oxford Handbook of Law, Regulation and Technology* (OUP, 2017).

[133] See, for example, *Report of the 2015 Informal Meeting of Experts on LAWS* (2 June 2015) UN Doc. CCW/MSP/2015/3, ¶¶ 39 and 51(a)(iv) <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G15/111/60/PDF/G1511160.pdf?OpenElement> accessed 24 June 2018 ("Several [States] expressed scepticism over the utility of [MHC], assessing it as being too vague, subjective and unclear…[MHC] may be useful as a policy approach to address shortcomings in current technology. However, it should not be applied as a legal criterion as this could undermine existing targeting law by introducing ambiguity").

> Third, there must be a means for the rapid suspension or abortion of the attack.[134]

Horowitz and Scharre criticise this as being an "idealized version of human control divorced from the reality of warfare and the weapons that have long been considered acceptable".[135] Certainly, on a literal interpretation this definition would prohibit most fire-and-forget munitions that remain in flight for any appreciable length of time, and this includes a large number of precision-guided munitions (PGMs). Yet, PGMs have existed for over 70 years and are now used by nearly every modern military,[136] thus their use is arguably lawful under customary international law.[137] Moreover, since their adoption and growth, PGMs have been instrumental in *reducing* civilian casualties,[138] as well as public and political tolerance towards such casualties.[139] Indeed, the substitution of PGMs for unguided munitions in an urban area may even turn what would otherwise be a war crime into a legitimate attack.[140] Arguably, the *ICRAC* definition would – if codified or seen as a distinct concept – at least lead to uncertainty. At worst, it would encourage a supplanting of less precise weapons merely by reason of having a human in-the-loop for longer, and this may needlessly sacrifice combatant and civilian lives.[141] A similar argument applies in relation to purely defensive systems, such as the *Phalanx*, which are effective precisely because they select and engage incoming threats before any human even knows of the latter's existence.[142]

---

[134] Frank Sauer, 'ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting', *ICRAC News* (14 May 2014) <https://www.icrac.net/icrac-statement-on-technical-issues-to-the-2014-un-ccw-expert-meeting/> accessed 24 June 2018.

[135] Horowitz and Scharre (n 109), 9.

[136] Barry D. Watts, *Six Decades of Guided Munitions and Battle Networks: Progress and Prospects* (CSBA, March 2007) <http://csbaonline.org/publications/2007/03/six-decades-of-guided-munitions-and-battle-networks-progress-and-prospects/> accessed 3 July 2018.

[137] See the examples of State practice in CIHL, Rule 17 <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule17> accessed 3 July 2018.

[138] Michael C. Horowitz and Paul Scharre, 'Do Killer Robots Save Lives?' *Politico Magazine* (19 November 2014) <http://www.politico.com/magazine/story/2014/11/killer-robots-save-lives-113010> accessed 3 July 2018 (noting that in World War II, air-dropped bombs only had a 50% chance of landing inside a 1.25-mile diameter. By Vietnam, this reduced to 50% falling inside an 800-foot diameter; now the diameter is as little as five feet. Consequently, smaller warheads and fewer munitions are required per target-hit, hence less collateral damage).

[139] Hugh White, 'Civilian Immunity in the Precision-Guidance Age' in Igor Primoratz (ed.), *Civilian Immunity in War* (OUP, 2010).

[140] See Human Rights Watch, 'Ukraine: Unguided Rockets Killing Civilians', *Human Rights Watch News* (24 July 2014) <https://www.hrw.org/news/2014/07/24/ukraine-unguided-rockets-killing-civilians> accessed 3 July 2018 (commenting on the indiscriminate use of unguided Grad rockets in Donetsk and how that may have amounted to a war crime, if done intentionally or recklessly).

[141] Crootof (n 130), 62 (comparing the use of PGMs with the deployment of a human-piloted bomber).

[142] Ibid., 61. This would breach the second limb of the *ICRAC* definition.

Accordingly, MHC should not be seen as a distinct legal concept, lest it dominates the legal analysis and blurs the clarity of existing rules.[143] Instead, it usefully draws attention to relevant considerations when applying the law to autonomous weapons, and it helps to clarify the limits to be prescribed on such systems. Thus, MHC may *facilitate the application* of existing rules *as they currently stand*; for example, in a LOAC Manual.[144] As will be seen through Chapters 5-7, the human factor is either already implicit in how LOAC compliance is secured,[145] or it is one of several ways to ensure such compliance.[146]

Moreover, as the above has demonstrated, MHC is open to variable criteria and even conflicting interpretations, which render it inherently imprecise.[147] Evolving these into a discrete legal standard, as opposed to encouraging their contextual application, risks blurring the clarity of existing laws.[148] Much worse, should this imprecision lead to human action supplanting machines in areas where automatic processing is more effective[149] – as noted in the PGM and *Phalanx* examples, above – this may risk both combatant and civilian lives, hence undermining LOAC norms.[150] Thus, Anderson and Waxman argue:

> [A]lthough some of its proponents view the MHC standard as flowing from LOAC, in some important respects it is quite at odds with the fundamental structure of LOAC, and its core principles of necessity, distinction, proportionality, and humanity.[151]

---

[143] Ibid.; Marauhn (n 132).

[144] To reiterate: MHC provides a suitable lens through which to evaluate how a LAWS deployment will meet LOAC norms, but this does not mean it should become the primary object of such a deployment.

[145] Marauhn, (n 132), 216-17 (discussing the principle of proportionality, where the need for nuanced and contextual assessment mean human judgment and control emerges from within the norm, not from the MHC concept *per se*. Hence, human control is not a legal requirement in itself, but an approach to ensure compliance).

[146] Ibid., 214-15 (discussing the principle of distinction and target verification, where in many scenarios involving large military objects by nature, either human-led or automated means of target identification may suffice).

[147] Anderson and Waxman (n 132), 1113-14 (discussing the "strategic ambiguity" of the concept, and the broad interpretation given to it by the US *versus* the narrow interpretation adopted by States allied with NGOs).

[148] Ibid; Marauhn (n 132).

[149] See 4.2.2 and 4.2.3 on the need for machine assistance and restrictions on human action.

[150] Crootof (130), 61-62.

[151] Anderson and Waxman (n 132), 1114.

The authors argue that the rules and principles of LOAC are concerned with intended or estimated *effects* on the battlefield, whereas MHC as presented in some definitions focuses on a certain *mode* of weapons and attack. Anderson and Waxman continue:

> It is not a law of nature…that weapons that put a human being 'meaningfully' in control of it, in some fashion, necessarily do the best job at minimizing battlefield harms. It is not beyond possibility that at some point, in some circumstances, a machine might do it better, on its own.[152]

To reiterate, the MHC concept is useful in that it draws attention to various LAWS-relevant issues. However, those issues are arguably better addressed with a contextual application of LOAC as it stands,[153] with the established content of the current rules and principles serving as an "interpretive floor".[154] Namely, priority must be accorded to the correct application of distinction, proportionality and precautions in attack, with MHC considerations serving to enhance the former's application. Any approach to MHC that prioritises human involvement in the use of force at the expense of human lives is inimical to the aims of LOAC and must, therefore, be rejected.

4.6 Conclusion

This chapter has demonstrated the practical need for human judgment and control in autonomy, as well as the legal requirement for this in the use of force. Consequently, it is difficult to imagine any use of LAWS that operates beyond human control complying with LOAC. However, this does not mean MHC should be codified or even treated as a distinct concept, as this may distort the crucial MN-H balance and emasculate the law. Instead, MHC should serve as an aid to the interpretation and application of LOAC in the context of LAWS; priority should be accorded to the substance of the LOAC rules and principles, which can act as an 'interpretive floor'. States and militaries should therefore identify the elements of MHC, as noted in this chapter, and integrate these into their national military manuals, training and the legal advice they provide to commanders. Of course, these same elements can also inform the drafting of a LOAC Manual by a group of experts.

---

[152] Ibid.

[153] Marauhn (n 132).

[154] Crootof (n 130).

A crucial point that has come out of this chapter is that LAWS will have an extensive life-cycle, with potentially seven human-machine interaction touchpoints. Thus, contrary to much academic analyses on LAWS, the focus of MHC should not be narrowly confined to force execution, where the system selects and engages targets. Instead, MHC may permeate the entire life-cycle in such a way that 'narrow loop' autonomy will potentially become less problematic. Perhaps the clearest example of such wider human control acting as a check on narrow loop autonomy – at least in a US/NATO context – is the Joint Targeting process.

Chapter 5
US and NATO Joint Targeting Doctrine as an Expression of Meaningful Human Control

5.1 Introduction

The previous chapter argued that there is a structural legal requirement to exercise human judgment and control over each 'individual attack'. Thus, inadequately restrained deployments of lethal autonomous weapon systems (LAWS) will *a priori* violate international humanitarian law (IHL)/law of armed conflict (LOAC). More specifically, there is a concern that if LAWS deployments do not benefit from a meaningful human control (MHC), the weapons may 'run amok' on the battlefield, and attack protected persons and objects in ways that will amount to more substantive violations of the LOAC rules. While many States and other commentators acknowledge that human-machine interaction occurs both upstream and downstream, those advocating a LAWS ban (or codification of MHC) tend to focus squarely on the *weapon* itself, and its critical functions at the point of weapons release. However, as will be seen in this chapter, such a narrow focus is inadequate on two grounds. First, it erroneously considers targeting to be a mere *action*, and fails to account for the elaborate (downstream) targeting *process* in which human judgment and control are pervasive.[1] Second, by focusing so narrowly on autonomy at the point of weapons release, LAWS critics lose sight of other (non-weapon) autonomous technologies, which apply no direct violent effects but are still highly influential in the critical function of target selection.

Accordingly, the following chapter contains four main parts.

- First, 5.2 lays down some key targeting definitions and target categories. Here, it will be seen that prioritised targets are of two types: a) specific targets, and b) targets belonging to a broader set. This crucial distinction affects how a LAWS will operate, the type of engagement, and the type and degree of human control in autonomous attack; yet most LAWS legal literature tends to ignore it.

[1] Hence, most literature that discusses LAWS 'targeting' are actually referring to the technical phenomenon of target *recognition*, as opposed to the human-led and deliberative targeting *process*.

- Second, 5.3 takes a more detailed look at the US and North Atlantic Treaty Organisation (NATO) targeting cycles, where it will be seen that human judgment and control has long been an integral feature of operational planning. The implication is two-fold. First, 'critical function' should be understood not just as a technical process at the point of attack, but as a highly deliberative human-led process, with multiple checks and balances at key points throughout the decisional pathway. Second, US and NATO forces are – save for some necessary adjustments – well-placed to deploy and use LAWS in a lawful manner.
- This is complemented by 5.4, which examines the 'central decision' in targeting; that is, the point of 'trigger-pull', after which humans can no longer influence the direct violent effects. Yet, before reaching this point, an entire operational planning process will have been undertaken, which can be expected to put the weapons operator in an informed position whether or not to launch the system/authorise further engagements, or cancel/suspend and re-plan the attack.
- Finally, 5.5 directly applies the MHC concept to the US/NATO targeting cycles. Here, two contrasting arguments are put forward. On the one hand, as a strongly human-centred processes, US/NATO targeting doctrine is likely to afford a high level of MHC to LAWS deployments; so long as States and their militaries resist the temptation to broaden the notion of an 'individual attack'. On the other hand, there is a risk that deliberative human control may be undermined by emerging autonomous technologies *within* the targeting process. Namely, the greater use of algorithms in military intelligence and target development may supplant human judgment in ways that will undermine MHC in LAWS (and other) deployments. Much will depend on how well such wider loop autonomy is tested, evaluated and integrated into the process.

5.2 Some Key Targeting Definitions and Categories

5.2.1 Target

In NATO military doctrine, a 'target' is:

> [A]n area, structure, object, person or group of people against which lethal and non-lethal capability can be employed to create specific psychological or physical effects.[2]

US military doctrine uses a similar definition, though with an explicit requirement of incurring a loss to the adversary.[3] Unquestionably, all targets must in the first instance be valid military objectives under Additional Protocol I[4] or its customary equivalent;[5] be they objects,[6] traditional combatants[7], or civilians taking a direct part in hostilities.[8] Moreover, in all cases, targets should be attacked in support of the commander's intent, objectives and guidance.[9]

5.2.2 Targeting and the Targeting Process

In both NATO and US doctrine, Joint Targeting[10] is:

> The process of selecting and prioritizing targets and matching the appropriate response to them, taking account of operational requirements and capabilities.[11]

This can be seen to have three key elements: a *process* orientation, an awareness of one's own *capabilities*, and a link back to specific *requirements*.[12] As above, this must logically begin with an objective, which drives the subsequent processes.[13] The

---

[2] North Atlantic Treaty Organisation (NATO), *AJP-3.9: Allied Joint Doctrine for Joint Targeting* (Edition A Version 1, NATO Standardisation Office, April 2016) (hereafter, AJP-3.9), 1-2.

[3] Joint Chiefs of Staff, *Joint Publication 3-60: Joint Targeting* (JCS, 31 January 2013) (hereafter, JP 3-60), I-1 ("A target is an entity (person, place, or thing) considered for possible engagement or action to alter or neutralize the function it performs for the adversary").

[4] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3 (hereafter, AP I).

[5] Restated in Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law, Vol. 1: Rules* (CUP, 2005) (hereafter, CIHL). All Rules available at: <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul> accessed 10 June 2018.

[6] Article 52(2), AP I; CIHL, Rule 8.

[7] Articles 43(2), AP I; CIHL, Rule 3.

[8] Article 51(3), AP I; CIHL, Rule 6.

[9] AJP-3.9, 2-2; JP 3-60, I-7. See also Geoffrey S. Corn and Gary P. Corn, 'The Law of Operational Targeting: Viewing the LOAC Through an Operational Lens' (2012) 47 Texas International Law Journal 337, 349.

[10] 'Joint' in this context refers to the joint effort between all components of the armed forces: Army, Air Force and Navy.

[11] AJP-3.9, LEX-7; JP 3-60, I-1.

[12] Phillip R. Pratzner, 'The Current Targeting Process', in Paul AL. Ducheine, Michael N. Schmitt and Frans PB. Osinga (eds.), *Targeting: The Challenges of Modern Warfare* (TMC Asser Press, 2016), 79.

[13] Ibid.

specifics of these processes are detailed in 5.3, and are informative on the deliberative nature of US/NATO targeting.

In practice, the Joint Targeting process is split into two categories:

- **Deliberate Targeting**, which focuses on future, scheduled plans, usually on a 24-72-hour time horizon.[14]
- **Dynamic Targeting**, for targets identified too late for deliberate targeting, and which are employed in current operational planning (usually the current 24-hour period).[15]

The time horizons are just one of the differences between deliberate and dynamic targeting. Pratzner points out that both are underpinned by four distinct steps: objectives and guidance, planning, execution, and assessment.[16] However, each targeting process applies a different set of specific phases or steps, to derive its own Joint Targeting Cycle; again, the details of each one are explained further below, in 5.3. There, it will be seen that military deployments typically involve a strong element of human decision-making. The result is arguably that machine discretion is largely bounded in favour of deliberative human control, and is restricted to (automatic processing) competences in which machines typically outperform humans.

5.2.3 Levels of Warfare and Command

There are three levels of warfare to which the above targeting cycles and their respective command structures relate.

- **Strategic level**: where higher *national* or *multinational* strategic security objectives and guidance are determined, then national resources are developed and used to achieve those objectives.[17]
- **Operational level**: where *campaigns* and *major operations* are planned and conducted, deploying tactical forces to achieve strategic objectives within specific theatres.[18]

---

[14] AJP-3.9, 1-2; JP 3-60, II-1. See 5.3.1.1-5.3.1.6, for detailed steps of the deliberate cycle.

[15] AJP-3.9, 1-3; JP 3-60, II-2. See 5.3.2.1, for detailed steps of the dynamic cycle.

[16] Pratzner (n 12), 80-87. See also Corn and Corn (n 9), 351.

[17] See *DoD Dictionary of Military and Associated Terms* (JCS, August 2018), 219, <http://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf> accessed 30 August 2018.

[18] Ibid., 173.

- **Tactical level**: where *individual battles and engagements* are planned and executed to achieve military objectives assigned to tactical units or task forces.[19] Those operating on the battlefield are said to be at the 'tactical edge'.

These progress from the broad and general, to the focused and specific; from higher-level political officials right down to individual soldiers in combat. As will be seen below, the strategic/operational/tactical distinction is crucial for allocating roles and responsibilities within the targeting process,[20] and it impacts the utility of proportionality assessments – both in general[21] and in a LAWS context.[22]

5.2.4 The Joint Targeting Cycles and Target Categories

First, under the 'deliberate' process, there are **planned targets**, which are subdivided into those that are *scheduled* or *on-call*.[23] 'Scheduled' targets are engaged at a specific time, which will often be selected to maximise the military advantage of the attack.[24] 'On-call' targets have actions planned, but not for a specific delivery time; the commander expects to locate these targets in sufficient time to execute planned actions.[25]

Second, under the 'dynamic' process, there are **targets of opportunity**, which are subdivided into *unplanned* and *unanticipated* targets.[26] 'Unplanned targets' are known targets that were included on the target list, but not initially selected for engagement, for various reasons.[27] Subsequent changes in target status (priority, access, or permission) could result in the need or opportunity to engage such targets during the current cycle.[28] 'Unanticipated targets' are unknown or not expected to be present in

[19] Ibid., 226.

[20] See 5.3.

[21] See 7.2.1.

[22] See 7.2.3.4.

[23] AJP-3.9, 1-2; JP 3-60, II-2.

[24] Ibid.

[25] Ibid.

[26] JP 3-60, II-2–II-3. Similarly, AJP-3.9 links dynamic targeting to 'unexpected targets', at 1-3. It also recognises 'time-sensitive targets' at I-2, which may be "fleeting targets of opportunity".

[27] JP 3-60, II-3 (noting that they may not have been nominated; were nominated but did not make the 'Joint Integrated Prioritized Target List'; or they were not expected to be available within the current targeting cycle).

[28] Ibid.

the operational environment; thus, they are not included on the target list and an evaluation of the target is needed, to determine engagement requirements and timing.[29]

In addition, the dynamic process is used to prosecute **Time-Sensitive Targets** (TSTs), which merit prioritisation and special consideration. These are specific targets designated either at the political level, or at the operational level by the Joint Force Commander (JFC), which:

> Require[e] an *immediate* response because they *pose* (or *will soon pose*) a danger to friendly forces or are highly lucrative, fleeting targets of opportunity whose successful engagement is of high priority to achieve campaign or operational objectives.[30]

Accordingly, the JFC will either allocate new intelligence collection and engagement assets to the TST, or will divert assets away from other (lower priority) targets, in order to subject it to the dynamic targeting process.[31]

These target types and their categorisation are summarised in Figure 5.1, below.

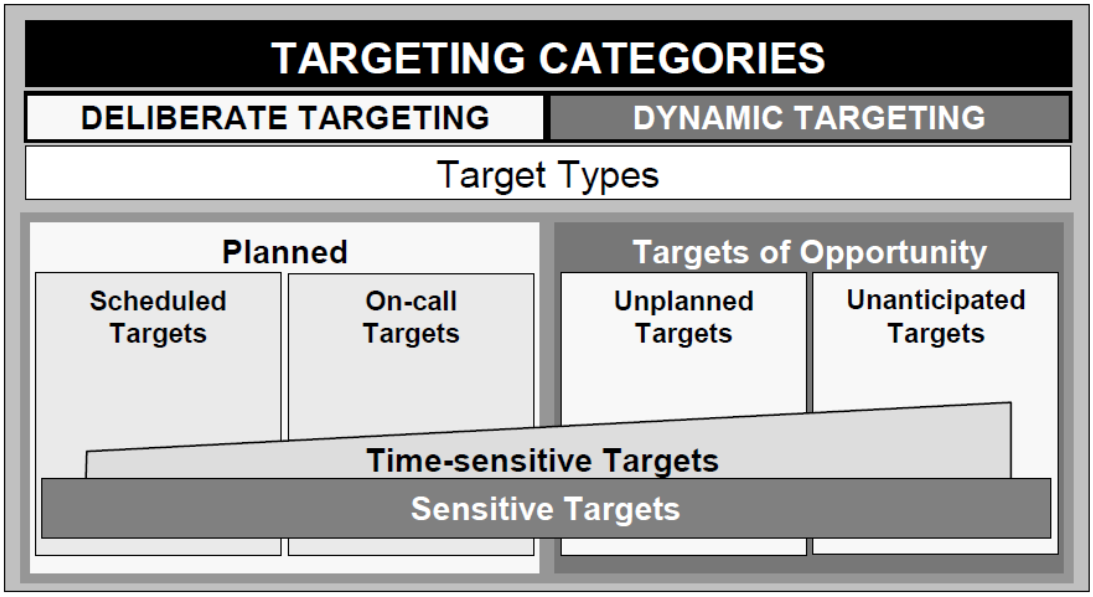

Figure 5.1: Targeting categories. Source: US Army, *Field Manual 3-60: The Targeting Process* (2010), 1-6.

[29] Ibid.

[30] AJP-3.9, 1-2.

[31] JP 3-60, I-9.

Three more types of targets should be mentioned.

- **High-Value Targets** (HVT) are targets that an enemy commander requires for the successful completion of his mission.[32] Accordingly, HVTs are determined by the value they offer to the *adversary*.[33]
- **High Pay-Off Targets** (HPT) are a subcategory of HVT, whose loss to the enemy will *significantly* contribute to the success of the friendly course of action.[34] HPTs are determined by the value they offer to *friendly* forces, rather than the adversary.[35]
- **High-Value Individuals** (HVI) are a subcategory of HVT, which consist of enemy personnel such as members of its leadership structure; or other personnel with particularly rare skills that are of high value to the adversary, such as bomb-makers.

HVTs and HPTs do not necessarily lean towards one or the other targeting cycle, and are potentially subject to either. HVIs, on the other hand, have tended to be prosecuted via the dynamic cycle[36] – albeit with extensive pre-strike deliberation[37] – because of their mobility and unpredictability as to when they might appear.

All of the above are **specific targets**, which are *individually* prioritised for engagement by the JFC or the political leadership (e.g. bridge X at location Y).[38] In contradistinction are **targets belonging to a set** (e.g. 'tanks' or 'artillery pieces'), which *as a whole* has been selected for attack, but within which specific targets have not been ascertained in advance, and will only be 'selected' during tactical combat.[39] This latter kind of engagement will see the dynamic targeting cycle largely autonomised and compressed, with deliberative human input occurring at a relatively high level of abstraction.

---

[32] Ibid. AJP-3.9, 1-2.

[33] AJP-3.9, 1-2.

[34] JP 3-60, I-9; AJP-3.9, 1-2.

[35] AJP-3.9, 1-2.

[36] Pratzner (n 12), 82.

[37] Ibid.

[38] Mark Roorda, 'NATO's Targeting Process: Ensuring Human Control Over (and Lawful Use of) 'Autonomous' Weapons' in Andrew P. Williams, and Paul Scharre (eds.), *Autonomous Systems: Issues for Defence Policymakers* (Headquarters Supreme Allied Commander Transformation, 2015), 158.

[39] Ibid.

5.2.5 Engagement Categories

With the above target categories in mind, there are three types of engagements that a LAWS may undertake.

5.2.5.1 Targeted Strike

In a LAWS context, a 'targeted strike' is broadly defined as:

> Any engagement by a LAWS that goes through a human-led deliberate or dynamic targeting cycle, to engage a specific and unique target.[40]

Such targets are usually HVTs, be they *planned targets* prosecuted via the deliberate process, or *targets of opportunity* under the dynamic process. Or, they may be extensively deliberated HVIs prosecuted under the dynamic process.

Hence, targeted strikes often involve attacks on fixed objects and pieces of infrastructure, like bridges and tunnels; and attacks on lucrative moving targets, such as warships. They also include attacks on HVIs that may suddenly appear during a mission,[41] though the relatively long time-horizon for biometric target recognition[42] will exclude autonomous targeted killing from the scope of this thesis.

When undertaking a targeted strike, a LAWS will arguably be analogous to a cruise missile,[43] but with the benefit of longer loitering times and technical discretion.[44] This enables it to vary the munition and blast radius (in the case of a platform) and the exact timing of attack, in accordance with perceptible circumstances on the ground. This will potentially enable those deploying and operating the system to meet certain precautionary obligations.[45]

---

[40] Maziar Homayounnejad, 'The Lawful Use of Autonomous Weapon Systems for Targeted Strikes (Part 1): Concepts, Advantages and Technologies', *TLI Think! Paper 11/2018* (2018), 10 <https://ssrn.com/abstract=3158170> accessed 7 July 2018.

[41] See Gregory S. McNeal, 'Targeted Killing and Accountability' (2014) 102 The Georgetown Law Journal 681, especially 701-730. (on advance deliberations and the development of kill lists).

[42] Ami Rojkes Dombe, 'Biometric Target Recognition', *Israel Defense* (16 March 2017) <http://www.israeldefense.co.il/en/node/28881> accessed 7 July 2018.

[43] 'Explained: How Cruise Missiles Work', *DefenCyclopedia* (1 August 2014) <https://defencyclopedia.com/2014/08/01/explained-how-cruise-missiles-work/> accessed 7 July 2018.

[44] See 2.3.4.1.

[45] See 7.3.

5.2.5.2 Tactical-Level Combat

This is where units undertake combat engagement (or pre-empt enemy attack) based on broad target categories that are approved during the deliberate cycle, and listed in mission Rules of Engagement (ROE); however, *they do not require approval for individual attack*.[46]

In contradistinction to targeted strikes, tactical-level combat involves attacks on numerous specific targets that *belong to a broader set*. Examples include 'uniformed enemy combatants', or routine military objects such as 'tanks', 'attack helicopters', or 'artillery pieces'. In a LAWS context, every such engagement will conform to generalised parameters programmed into the system's control software, and stored in its target identification library.[47]

When undertaking tactical-level combat, a LAWS will arguably be analogous to a more sophisticated sensor-fused weapon,[48] though it will operate with far greater levels of autonomy and discretion, a significantly longer time-of-flight, and multiple target sets, munitions and blast radiuses (in the case of a platform). Again, some of these additional technical features are useful devices for discharging the JFC's/weapons operator's precautionary obligations.

5.2.5.3 Platform-Defence

A third kind of engagement is when a LAWS comes under attack whilst on deployment and responds in self-defence or, more precisely, 'platform-defence'. While it is not yet certain that deploying militaries will authorise their systems to do this,[49] and there may be good reasons *not* to allow it,[50] platform-defence should still be considered in view of the tactical and operational imperative to preserve combat capability. This will be further discussed at 5.3.2.2.

---

[46] Roorda (n 38), 158.

[47] See 2.5.4.1 on standard ATR approaches.

[48] 'CBU-105 Sensor Fuzed Weapon: USAF's Ultimate Tank-Buster', *DefenCyclopedia* (12 June 2015) <https://defencyclopedia.com/2015/06/12/cbu-105-sensor-fuzed-weapon-usafs-ultimate-tank-buster/> accessed 7 July 2018 (providing a step-by-step account of the operation of the US Sensor-Fuzed Weapon).

[49] Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (Norton, 2018), 328.

[50] Ibid. (noting the risk of fratricide, unintended escalation of a crisis, and civilian/human shield casualties).

The engagement categories are summarised in Figure 5.2, below, and linked to their various features and characteristics that will be discussed throughout this chapter.

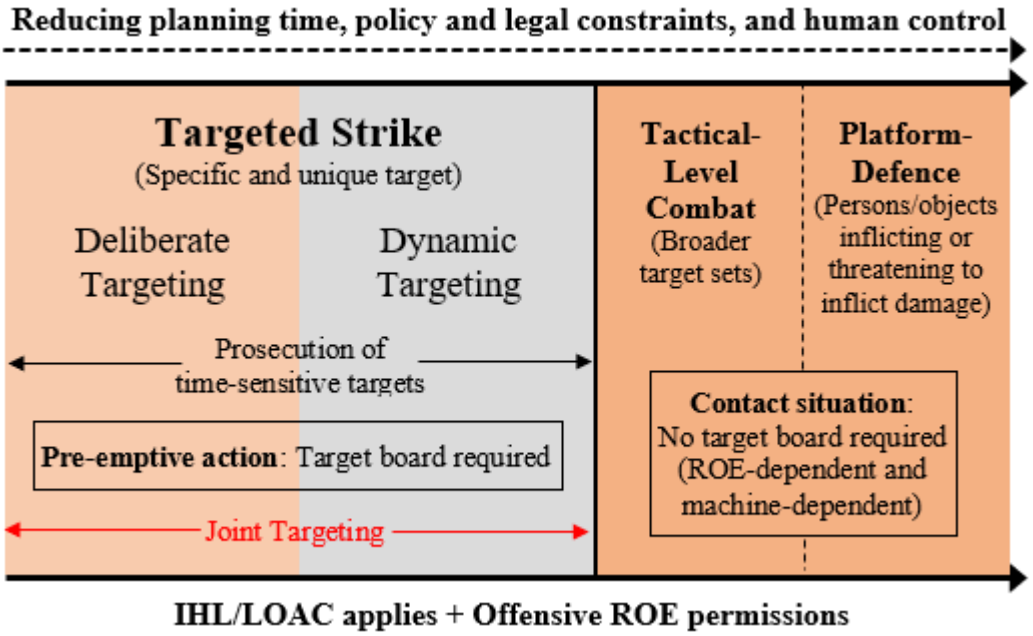

Figure 5.2: The engagement continuum. Source: Adapted from Figure 1.1, AJP-3.9, 1-4.

5.3 A Closer Look at the Joint Targeting Cycles

Turning now to a more detailed look at the Joint Targeting Cycles, the following will describe these with application to situations where the deployment of LAWS may be an option.

5.3.1 The Deliberate Joint Targeting Cycle

Before the deliberate targeting process formally begins, there is first **political direction** from the President and the Secretary of Defense (US), or the North Atlantic Council (NATO), by way of a) strategic military goals issued to the Joint Force Commander (JFC), b) guidance on campaign execution, c) approved target sets, including possible priority/time-sensitive targets (TSTs), and d) guidance on how specific targets within the broader sets will be selected for attack.[51] The JFC then formulates **operational-level goals** with which all subsequent (tactical) activities must conform. If he wishes to appoint target sets that have not been approved at the political

[51] AJP-3.9, 3-1, B-1 (listing examples of potential target sets). See also Roorda (n 38), 155.

level, explicit approval must be sought.[52] Accordingly, while the strategic goals issued at the political level are often very general and the target sets relatively broad, it is clear that target selection is largely human-controlled, from top to bottom.[53]

Once the JFC begins formulating operational goals, the deliberate targeting cycle formally kicks off. In both US and NATO doctrine, this contains six specific phases. As can be seen in Figures 2.3 and 2.4 below, each of the methodologies is cyclical and iterative, and there are substantial similarities between them, with only slight differences in wording and emphasis.

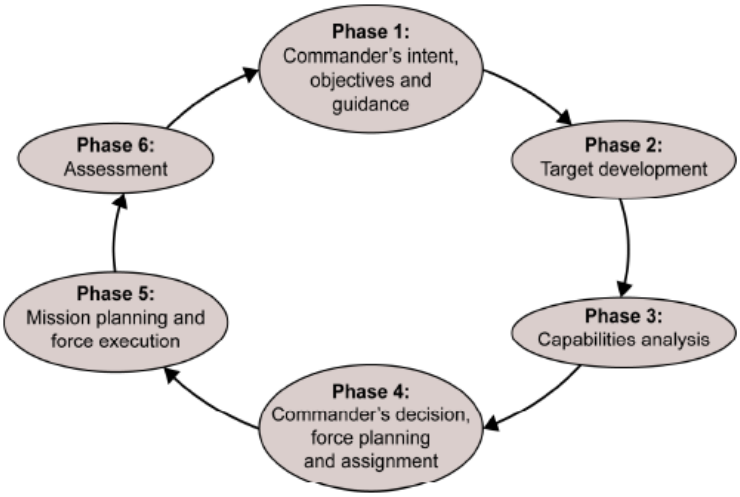

Figure 5.3: The NATO Joint Targeting Cycle. Source: AJP-3.9, 2-2.

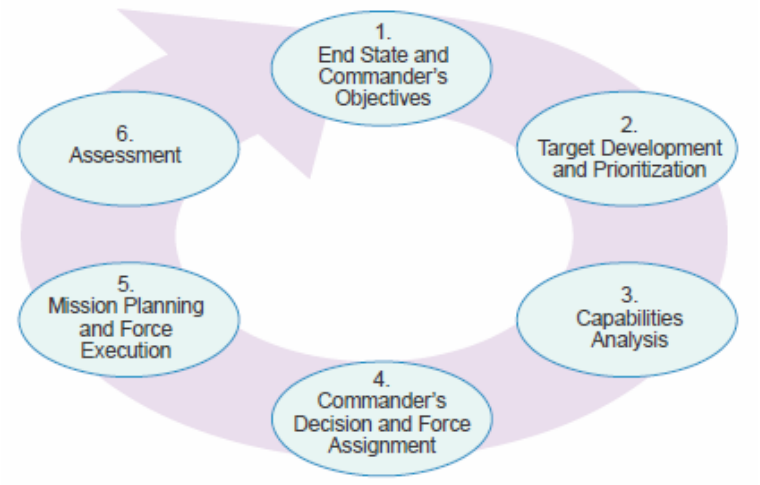

Figure 5.4: The US Joint Targeting Cycle. Source: JP 3-60, II-4.

[52] Roorda, ibid., 156; AJP-3.9, B-1 (noting the particular importance of verifying the military status of any civilian objects in accordance with LOAC, before seeking such specific approval).

[53] Even more so with politically-approved TSTs and HVIs.

5.3.1.1 Phase 1: (End State and) Commander's Intent, Objectives and Guidance

Both cycles begin with the **commander's intent, objectives and guidance**, which aim to achieve the military **end state**; that is, the desired set of conditions needed to resolve the situation or conflict at hand. As mentioned above, this is where political and strategic guidance from higher headquarters is methodically translated into an operational plan.[54] JFCs clearly identify a) what to accomplish, b) under what circumstances, and c) within which parameters. Targeting staffs will have to understand these goals, guidance and intents in relation to politically-approved target sets, in order to translate them into desired effects and concrete tasks that are logically related to the end state.[55] As emphasised in US doctrine, "[a]ttainment of clear, measurable, and achievable objectives is essential to the successful realization of the desired end state".[56] The ability to fulfil this with the correct type and extent of combat power is the hallmark of effective targeting,[57] and this may be an argument for including LAWS in military arsenals, as they afford commanders the *option* of autonomous lethal targeting where this is judged to be an objectively superior means to achieve the desired end state.

Moreover, these early demands for clear, measurable and achievable objectives are *necessary* (though not always sufficient) where the use of software-controlled LAWS is contemplated at a later stage. This is because robots can only undertake tasks that are specified in *precise* programmable detail, with *tangible* outcomes.[58] Clear, measurable and achievable objectives are a necessary starting point for setting objective and programmable parameters.

5.3.1.2 Phase 2: Target Development (and Prioritisation)

In this second Phase, potential targets that can be attacked to achieve JFC objectives are **identified**, **developed**, **nominated** and **prioritised**. In US doctrine, this occurs in three stages (NATO carrying out broadly similar tasks, but grouped into five stages[59]):

---

[54] AJP-3.9, 2-2–2-3; JP 3-60, II-4–II-5.

[55] Roorda (n 38), 156.

[56] JP 3-60, II-4–II-5.

[57] Ibid., II-5.

[58] See 2.2.3.2.

[59] AJP-3.9, 2-3–2-4 (describing target analysis, vetting, validation, nomination, and prioritisation).

(1) **Target System Analysis** (TSA). This is an all-source (intelligence) examination of *functionally interrelated* targets,[60] which contribute to adversary capabilities.[61] The TSA is a vital starting point, as the real importance of a nominated target lies not only in its own characteristics, but also in its *relationship to other targets* within an operational system.[62] Thus, a TSA identifies critical vulnerabilities within the adversary's system which, if targeted by the appropriate capability, would achieve JFC objectives.[63] This is essential to isolate individual HVTs for a more efficient application of force, rather than attempting to attack and destroy an entire operational system.[64] Equally, the TSA identifies substitutability within the adversary's system, which enables the grouping of targets that must be destroyed or neutralised together.[65]

(2) **Entity-Level Target Development**, which builds on the TSA by generating sets of essential data on each 'entity' (potential target) in three sequential stages: basic, intermediate and advanced.[66] This progresses an entity from initial identification and functional characterisation, through to execution-level detail; at which point the target is considered fully developed.[67] Once a target is nominated for development, an Electronic Target Folder (ETF) is started, which stores entity-level target intelligence, operational, planning, and legal information.[68] Once target development reaches 'intermediate' stage, the entity is placed on a Candidate Target List (CTL), which drives further target development and list management.[69]

(3) **Target List Management** (TLM). This third step begins when an entity is first nominated for development, and ends with a prioritised target list. Concretely, it

---

[60] JP 3-60, II-5 (describing a 'target system' as "a collection of assets directed to perform a specific function or series of functions" within an (adversary's) operational system. For example, the components of a 'ballistic missile target system' might include transporter erector launchers, resupply vehicles, command and control links and nodes, meteorological radars, missile fuel storage facilities, and the supporting transportation network).

[61] AJP-3.9, 2-3 and LEX-8; JP 3-60, II-5–II-11.

[62] JP 3-60, II-5.

[63] AJP-3.9, LEX-8.

[64] JP 3-60, II-9.

[65] For example, where two bridges cross a strategically significant river, destroying one may give no military advantage, whereas a coordinated attack on both would. Similar logic applies to alternative power sources, such as two electricity generators.

[66] JP 3-60, II-9.

[67] Ibid.

[68] Ibid.

[69] Ibid.

sees entities on the CTL undergoing *vetting* and *validation* processes, which lead to target *nomination* and *prioritisation*:

- **Target vetting** assesses the accuracy of the supporting intelligence, to verify that the candidate target performs the specified function for the adversary, and also to verify its significance within the operational system.[70]
- **Target validation** ensures all vetted targets continue to meet the JFC's objectives, guidance and intent; that they comply with both LOAC norms and mission ROE;[71] and that they are coordinated and deconflicted.[72]

All vetting session votes are recorded in the ETF.[73] Throughout these processes, two distinct sets of staffs work to refine and further specify the broader target sets:

- **Targeteers** will a) analyse adversary capabilities, b) determine the best targets to engage to achieve the designated goals, and c) using intelligence, surveillance and reconnaissance (ISR) assets, will collect essential information on the target to verify its battlefield status and to be able to engage it.[74] Early issues concerning potential collateral damage and other undesired/collateral effects[75] are also noted, for further analysis in Phase 3.
- Concurrently, **legal advisors** review every proposed target, to ensure compliance with both LOAC norms and mission ROE,[76] as well as any other specific restrictions such as 'no-strike lists'.[77]

---

[70] AJP-3.9, 2-3; JP 3-60, II-11.

[71] AJP-3.9, 2-3–2-4; JP 3-60, II-11. As Corn and Corn (n 9) explain at 353-57, LOAC sets *legal* limits for defining and engaging lawful targets, while ROE are an additional source of authority from within the political and military spheres, which define guidelines for permissible combat action. Thus, while ROE must be consistent with LOAC, they are technically not law, but are statements of *policy* based partly on LOAC, and further restrained by diplomatic and political factors, and operational requirements.

[72] Coordination with other agencies and operations is necessary to avoid collision, fratricide and collateral damage.

[73] JP 3-60, II-11.

[74] Roorda (n 38), 157.

[75] While neither 'collateral damage' nor 'collateral effects' are legal terms, they are used interchangeably here as shorthand for both incidental damage to civilian objects, and incidental death and injury to civilians.

[76] Corn and Corn (n 9), 352.

[77] Ibid.; AJP-3.9, 2-3–2-4; JP 3-60, II-5–II-13.

Once targets are thoroughly vetted and validated, they are added to the Joint Target List (JTL) or the Restricted Target List (RTL).[78] From these, individual targets are prioritised in accordance with JFC objectives and nominated for the Joint Prioritised Target List (JPTL).[79] Often, this process benefits from added review and oversight from the Joint Targeting Coordination Board (JTCB).[80]

At this stage, it is worth reiterating that there are – conceptually and practically – two categories of targets that may emerge in Phase 2.

- **Specific targets**, which have been sufficiently developed and validated to be *individually* nominated for attack (e.g. bridge X at location Y).
- **Targets belonging to a set**, which *as a whole* has been selected for attack, but within which specific targets have not yet been sufficiently developed and validated (e.g. 'tanks', 'artillery pieces', or 'uniformed enemy combatant').

The former category essentially describes planned targets for which there are sufficiently detailed target data to directly *schedule* their engagement, or to hold them *on-call* to be prosecuted if the situation later demands it. Engaging this category is the very essence of a targeted strike in which a LAWS will behave like a more sophisticated cruise missile.

By contrast, targets belonging to a broader set lead to tactical-level combat and they do not require human approval for individual attack. Each target will be selected in accordance with generalised parameters programmed into the machine's control software. The LAWS will prosecute each one through an autonomised (and compressed) dynamic cycle,[81] which enables it to behave like a more sophisticated sensor-fused weapon.

---

[78] Corn and Corn, ibid.; JP 3-60, II-11. Both lists contain valid targets, the only difference being that those on the RTL are subject to specific restrictions (e.g. on attack timing), due to operational considerations.

[79] JP 3-60, II-12.

[80] AJP-3.9, 4-7–4-8; JP 3-60, III-3–III-5 (explaining that the JTCB is established by the JFC, and is comprised of representatives from the different battle staffs and components (land, air, sea, special operations), and sometimes other agencies and multinational partners. The Board synchronises and manages Joint Targeting efforts in accordance with JFC objectives at the operational level).

[81] See 5.3.2.1, for detailed steps of the dynamic cycle.

Of course, JFCs are not restricted to one or the other, and it is possible that a LAWS may be deployed on a multi-strike mission to engage a specific target *and* broader sets of targets.

5.3.1.3 Phase 3: Capabilities Analysis

Once targets have been vetted and validated, approved onto the JTL or RTL, and nominated for the JPTL, the JFC will undertake a detailed **capabilities analysis** in relation to the desired effects and objectives.[82] Collateral damage estimation (CDE), which began in the previous stage, remains a critical component of the analysis here.[83] Consequently, no mean or method of attack will be considered or launched where it is expected to cause 'excessive' collateral damage in relation to the concrete and direct military advantage anticipated,[84] and some may be discarded altogether if others would better avoid or minimise collateral damage.[85] In practice, the JFC and his battle staffs will begin with a rigorous focus on specific 'precautions in attack', rather than the amorphous notion of 'proportionality' and 'excessiveness' of collateral damage.[86] Once all such precautionary measures have been explored, it is more likely that proportionality compliance – highly elusive when attempted in the abstract – will be more easily secured.[87] Moreover, a mean or method will be rejected for particular targets if it is judged unlikely to achieve the desired effects.[88] Conversely, if there are no apparent concerns, the weapon will be included among the options for the JFC to decide upon.

Thus, it is during this Phase where the advantages of autonomous attack will be considered in more detail, to ensure that a LAWS is able to engage the target in a way that generates the desired effects, to achieve the objectives in accordance with operational and legal standards.[89] For example, will an autonomous drone destroy or neutralise the prioritised target, while maximising the employment efficiency of

---

[82] AJP-3.9, 2-4; JP 3-60, II-13–II-16.

[83] AJP-3.9, 2-4.

[84] Articles 51(5)(b) and 57(2)(iii), AP I; CIHL, Rule 14. See also 7.2.

[85] Article 57(2)(a)(ii), AP I; CIHL, Rule 17. See also 7.3.2.2.

[86] Geoffrey S. Corn, 'War, Law, and the Oft Overlooked Value of Process as a Precautionary Measure' (2015) 42 Pepperdine Law Review 419, 435-40. See also 7.3, especially 7.3.5.1.

[87] Ibid. See 7.3.5.1.

[88] Roorda (n 38), 159.

[89] AJP-3.9, 2-4; JP 3-60, II-13. See also 1.2.2 on the advantages of LAWS.

forces, and minimising collateral effects?[90] Specific, relevant characteristics that JFCs may consider include the superior data-processing capabilities and the greater responsiveness of machine action, relative to humans.[91] These may give a LAWS the upper hand in engaging an unanticipated but fleeting target, or in cancelling/suspending an attack within a split-second of civilians unexpectedly appearing on the scene.[92] Part of the analysis will also include mitigating features, such as the likely performance of software suppressors in the proposed operational environment.[93]

In addition, the capabilities analysis stage involves extensive weaponeering (for prioritised targets)[94] and consideration of non-lethal capabilities,[95] to proactively avoid or minimise the CDE.[96] An example of such a capability is the *CBU-107* Passive Attack Weapon, which dispenses non-explosive penetrator rods aimed at destroying 'soft' (non-armoured) targets, without causing incidental injury or damage to surrounding structures.[97] Non-lethality is becoming increasingly important in contemporary targeting practice, not just for civilian risk mitigation, but also to reduce opportunities for adversary propaganda and for cost reasons.[98] For a LAWS, this may require the use of a platform rather than a munition, as the former is potentially more amenable to delivering a range of capabilities, including non-lethal ones. Moreover, where non-lethality is desired in a relatively dangerous area, LAWS may be an overall better option because of the absence of mortal risk.[99]

---

[90] Ibid.

[91] See 1.2.2 and 4.2.

[92] See the Grdelica incident in 7.3.2.3, (notes and text accompanying) nn 204-209 therein; and the inadvertent bombing incidents in 6.5.3.4.5, n 323 therein.

[93] Ronald C. Arkin, *Governing Lethal Behaviour in Autonomous Robotics* (Chapman & Hall/CRC, 2009).

[94] JP 3-60, II-14–II-15. Weaponeering is the process of 'designing an attack' by selecting a specific weapon, tailoring the type and amount of warhead, and tailoring the fusing mechanism. The aim is to inflict the desired damage to the intended target, while minimising collateral effects.

[95] AJP-3.9, 2-4; JP 3-60, II-15–II-16.

[96] Article 57(2)(a)(ii), AP I; CIHL, Rule 17.

[97] See Human Rights Watch, *Off Target: The Conduct of the War and Civilian Casualties in Iraq* (Human Rights Watch, 2003), 46-49 (explaining how, in March 2003, US forces planning at attack on the Iraqi Ministry of Information switched from the *Hellfire* missile to the *CBU-107* 'rods from God' weapon. The aim was merely to destroy the antennae on the roof of the building, thereby denying any broadcast capability to the Iraqi Government, but without destroying the facility itself, or any of the nearby churches, mosques or civilian dwellings).

[98] Pratzner (n 12), 93 (noting that, when compared with complex weapon systems, non-lethal capabilities are "pennies on the dollar in comparison").

[99] Non-lethal munitions can also provide an effective precautionary measure, by way of advance warning under Article 57(2)(c), AP I, and customary norm restated in CIHL, Rule 20. See also 7.3.2.4.

5.3.1.4 Phase 4: Commander's Decision, Force Planning and Assignment

The fourth stage involves the **commander's decision, force planning and assignment**, where the outputs of the previous stage are integrated with any further operational considerations.[100] Namely, there is a fusion of the capabilities analysis with available forces; intelligence and ISR sensors; and weapon systems.[101]

Put simply, the JFC, often supported by a JTCB, will match capabilities against targets and assign those capabilities accordingly.[102] For example, consider the situation where multiple fixed HVTs must all be engaged simultaneously, yet there is also a shortage of aircrews. In such a case, a remotely-piloted drone would be relatively inefficient and ineffective,[103] thus possibly necessitating the deployment of multiple autonomous drones all overseen by a single operator. Therefore, a capability is matched to a selected target or targets, to achieve the desired effect within resource constraints. In practice, however, scarcity of resources may mean that targets are often selected based on the available weapons and their characteristics.[104]

Either way, matching a given capability to a given target is a conscious decision, and is taken with the benefit of additional human oversight and scrutiny from the JTCB. Once it is done, the JFC issues final approval for prioritised targets, thereby deriving the JPTL. Subsequently, tasking orders are prepared and released to executing components and forces.[105] Any relevant constraints, restraints and precautions that have emerged during these first four Phases are passed onto the assigned units, and they can be adapted to be more or less strict, depending on the unit's ability to ensure compliance with the relevant obligations.[106]

---

[100] AJP-3.9, 2-4.

[101] JP 3-60, II-16 (pointing out that this process links theoretical planning with actual operations).

[102] Roorda (n 38), 159.

[103] Merel Noorman, 'Responsibility Practices and Unmanned Military Technologies' (2014) 20 Science and Engineering Ethics 809, 818 (noting that it can take up to 168 people to keep an armed *Predator* drone airborne for 24 hours).

[104] Roorda (n 38), 159. Consider a situation where it is desired to kill or disrupt a given HVI and there are cruise missiles available, but no drones. Commanders may opt to destroy the HVI's compound instead, as this can be done via cruise missile, whereas a targeted killing will likely require a drone.

[105] JP 3-60, II-20.

[106] Roorda (n 38), 160.

Accordingly, the 'commander's decision' in this Phase is to approve the draft JPTL; to approve individual targets to be added/removed from it; and/or to approve a particular way of engaging a given target, if needed.[107]

5.3.1.5 Phase 5: Mission Planning and Force Execution

The fifth Phase is in two parts. First, **mission planning** (5a) is carried out by unit/component commanders who largely replicate Phases 1-4, but on a more detailed and tactical level.[108] Goals are re-evaluated, additional intelligence is collected, targets are further refined, and means and methods are chosen from within the assigned unit that are best suited to achieve the goals.[109] It is during this Phase that targeting staff obtain final positive identification (PID) and combat identification (CID) of targets, before engaging.[110] Accordingly, final precautionary measures are taken to:[111]

- verify that targets are in fact lawful military objectives;[112]
- ensure all feasible steps have been taken in the choice of means and method of attack, to avoid or minimise collateral damage;[113]
- provide effective advance warning to the civilian population, to further reduce collateral effects, unless circumstances do not permit;[114] and
- refrain from launching the attack if the expected collateral effects are 'excessive' in relation to the military advantage anticipated.[115]

Phase 5a planning to comply with the above legal norms will become very granular, and will include assessments of the:

> [L]ocation, type, size and material of target; civilian pattern of life; time of attack (day or night); weapon capabilities; weapon effects; direction of attack; munition fragmentation patterns; secondary explosions; infrastructural collateral concerns, personnel safety; and battlespace deconfliction measures.[116]

---

[107] JP 3-60, II-19–II-20.

[108] Roorda (n 38), 160.

[109] Ibid.

[110] For the detailed steps, see AJP-3.9, 2-5–2-6; JP 3-60, II-20–II-30; and 5.3.2.1.

[111] See 7.3.2 for more detail on the legal norms obligating these precautionary measures.

[112] Article 57(2)(a)(i), AP I; CIHL, Rule 16.

[113] Article 57(2)(a)(ii), AP I; CIHL, Rule 17.

[114] Article 57(2)(c), AP I; CIHL, Rule 20. An example of such circumstances may be where an element of surprise is needed to maximise the military utility of the attack.

[115] Article 57(2)(a)(iii), AP I; CIHL, Rule 19.

[116] Roorda (n 38), 160.

Subsequently, a unit commander will approve the operation, as well as the use of a particular mean or method – be that manned, remotely-piloted or autonomous – against a specific target or target set.

Moving into the **force execution** stage (5b), precautionary measures continue to be applied.[117] Indeed, key to success here is flexibility, as combat operations are inherently dynamic.[118] During force execution, the operational environment constantly changes because of actions taken by the Joint Force, by the adversary, and by other actors.[119] Thus, constant attention must be paid to PID, CID and target validation.[120] Should there be a change of circumstances that makes continuation of the attack unlawful, it must be cancelled or suspended.[121] Concretely, during this stage, both US and NATO forces utilise the F2T2EA methodology, or the 'kill chain' as it is often known. That is, Find, Fix, Track, Target, Engage and Assess.[122] This is effectively a 'targeting cycle within a targeting cycle', to accommodate the battlefield changes mentioned above. Importantly, it is identical to the dynamic targeting cycle, the individual components of which are detailed below.

One crucial moment during force execution, which will be further explained below, is worth flagging up here: the point at which a *human* decision is made, which leads to potentially irreversible violent action.[123] In most cases, this is the decision to fire or launch a weapon; for example, when a sniper pulls the trigger.[124] As the above clearly demonstrates, this vital 'decision point' comes after a long series of preparatory analysis and decisions by multiple other staffs and the JTCB. Yet, it is a crucial point of focus here, as it effectively is the point of 'no return'; the last opportunity to ensure deliberative human judgment in the application of lethal force, for effective accountability and adherence to operational and legal requirements.

---

[117] Ibid.

[118] AJP-3.9, 2-4.

[119] JP 3-60, II-20.

[120] Ibid., II-21.

[121] Article 57(2)(b), AP I; CIHL, Rule 19. See also 5.4 on 'The Central Decision in Targeting'.

[122] AJP-3.9, 2-5–2-6; JP 3-60, I-8 and II-21–II-30. NATO also includes 'Exploit' after 'Engage'.

[123] Roorda (n 38), 160.

[124] Ibid.

5.3.1.6 Phase 6: Assessment

The final stage is **assessment**, which seeks to measure if, and to what extent, the planned effects have been realised, after tactical activities have been executed.[125] This is done largely via battle-damage assessment (BDA), munitions effectiveness assessment (MEA) and collateral damage assessment (CDA).[126] On an operational level, this stage is vital to the iterative nature of the targeting process, as it contributes to wider campaign assessment and assists JFCs in future decision-making.

Overall, the US and NATO deliberate targeting cycles both represent detailed planning, fully informed by legal obligations and ROE constraints; careful alignment of targeting efforts with both internal capabilities, and broader political/strategic priorities; and post-attack assessment, to enable an iterative approach that brings about continuous improvement in targeting efforts. It is, without doubt, a *human*-centred process, which benefits from extensive professional judgment and a sequential approach to target development, nomination and prioritisation.

5.3.2 The Dynamic Joint Targeting Cycle

As mentioned above, the dynamic targeting cycle has a dual role: to engage **planned targets** at Phase 5b (the tactical stage) of the deliberate cycle; and to prosecute **targets of opportunity** (including TSTs), which are identified *during* a tactical mission, when it is too late for their inclusion in the deliberate cycle.

5.3.2.1 Dynamic Targeting Steps

The dynamic targeting cycle comprises the F2T2EA steps (the 'kill chain'), as follows:[127]

- **Find**: ISR assets (sensors) search for, and detect, potential targets that meet the initial criteria (as set by JFCs) in the designated areas. Once detected, potential targets trigger actions to determine whether further attention (Fix), or deviation from existing plans (as is the case with TSTs) is necessary. In the case of the latter, the result of a Find is a (time-sensitive) target nomination for further refinement, and a reallocation of resources.

[125] For full detailed steps, see AJP-3.9, 2-6–2-7; JP 3-60, II-31–II-36.

[126] Ibid.

[127] Summarised from AJP-3.9, 2-5–2-6.

- **Fix**: focused sensors allow staff to identify and geolocate the target via cross-cueing and intelligence-fusing. An initial risk assessment is also done on the target and, in the case of a TST, there is a determination of the target's window of vulnerability.
- **Track**: sensors are assigned and prioritised to track a target; to maintain contact and monitor the target as it moves through the environment. This continues until an engagement decision is finalised, and the next two steps are fully prosecuted.
- **Target**: all restrictions (including LOAC, ROE, no-strike lists and deconfliction) are satisfied; engagement capabilities are aligned with the desired effect; the risk assessment is complete; force packaging is complete (determining the weapon-to-target match); and final engagement approval is granted.
- **Engage**: target is struck with the approved weapon, and from the pre-determined attack geometry. Both the target and its engagement continue to be closely monitored, to maintain situational awareness.[128]
- **Assess**: the engagement is reviewed, to determine whether the desired effects have been created, and the objectives achieved. If not, then the outcome of this step may be a decision to re-engage, using either the same or a different capability. In the case of TSTs, HVTs and HPVs a rapid initial assessment is vital for any re-engagement to go ahead.

The above steps are used in manned or remotely-piloted targeting to prosecute *specific* targets (be they planned, or of opportunity), but *not* targets belonging to a broader set. The latter implicates tactical-level combat which does not need approval for individual attack. Conversely, a LAWS will be likely to autonomise (and compress) an adapted F2T2EA cycle when prosecuting any of its target categories, including those belonging to a broader set.[129]

[128] As noted above, there is a central 'decision point' here, after which irreversible violent action will take place. The implications of this for deliberative human control of a weapon system are further examined in 5.4.

[129] While this is not explicitly stated in any authoritative document, it is arguably a necessary assumption for a LAWS that will operate autonomously throughout the dynamic process.

5.3.2.2 Inverting the Dynamic Targeting Steps for 'Platform-Defence'

When an attacking LAWS is on a mission, it may be subject to a risk of attack by the adversary's weapon systems, which will also apply F2T2EA, or a similar approach derived from Pratzner's four common steps.[130] If so, then the former (attacking) LAWS will also need to defend itself, possibly using an inverse set of capabilities. Howis suggests HELD2O, which is Hide (to counter the adversary's Find capability), Evade (Fix), Lose/Manoeuvre (Track), Deceive (Target), Defend (Engage) and Obscure (Assess).[131] While the exact details of this 'defensive chain' may be subject to some debate and will certainly need developing, the main point is indisputable: even without presenting any direct mortal risk to attacking forces, a LAWS will need these defensive capabilities to preserve friendly combat capability.

The importance of this point should not be underestimated: preserving combat capability is an integral feature of US doctrine on Joint Operations,[132] and is a distinct tactical and operational imperative.[133] It is also a "central component of mission accomplishment",[134] and is indispensable for bringing about the prompt submission of the enemy. Not only is this a clear military advantage, it is also consistent with civilian risk mitigation,[135] as 'mission accomplishment' may bring forward a complete end to armed conflict. As a corollary, preserving combat capability is focused on protecting *operational capacity*, not just the lives of friendly forces.[136] Thus, the concept undoubtedly applies to autonomous systems, which will be expensive items in limited supply and vulnerable to seizure;[137] notwithstanding the absence of mortal risk when these are snatched or destroyed in combat. On a practical note, effectuating the

---

[130] Pratzner (n 12), 80-87.

[131] Comments of Digby Howis, 18 December 2017, at: <https://twitter.com/DigbyHowis/status/942861600356155392> accessed 7 July 2018.

[132] JP 3-0, III-35–III-42 (detailing the three elements of the *Protection* Joint Function, of which 'force protection' is the most relevant, as it incorporates the preservation of combat capability).

[133] Ibid., VIII-18 (stating that in seizing the initiative, "JFCs *must strive to conserve the fighting potential* of the joint/multinational force at the onset of combat operations") (emphasis added).

[134] Geoffrey S. Corn and James A. Schoettler, 'Targeting and Civilian Risk Mitigation: The Essential Role of Precautionary Measures' (2015) 223 Military Law Review 785, 824-25.

[135] Ibid., 825-26; US Army, *Army Doctrine Reference Publication 3-0: Unified Land Operations* (US Department of the Army, May 2012), ¶ 1-7 (linking the attainment of "operational advantages" in combat to "destroying the enemy with *minimal losses to friendly forces* as well as *civilians and their property*") (emphasis added).

[136] Corn and Schoettler, ibid., 826.

[137] Emily Tamkin and Paul McLeary, 'China Seizes US Navy Drone in South China Sea', *Foreign Policy* (16 December 2016) <https://foreignpolicy.com/2016/12/16/china-seizes-u-s-navy-drone-in-in-south-china-sea-raising-stakes-president-trump/> accessed 7 July 2018.

preservation of combat capability may be done through *active* or *passive* means,[138] and it is arguably for JFCs to determine the best combination of defensive actions.

With this in mind, most of the HELD2O 'defensive chain' – Hide, Evade, Lose, Deceive –consists of passive defensive measures, analogous to those mentioned in JP 3-0, such as camouflage, concealment, deception, and dispersion.[139] A particularly effective measure for an autonomous system would be to pull off "genuinely random and unpredictable manoeuvres",[140] which an untethered LAWS – unencumbered by GPS latency, and endowed with the speed and precision of automatic processing – may carry out to confuse an opponent challenging it with direct fires. Such passive capabilities are likely to be uncontroversial, as they aim to *avoid* kinetic attack. However, to ensure its effective preservation, a LAWS may also need *active* defensive capabilities; that is, "direct defensive actions…to destroy, nullify, or reduce the effectiveness of hostile air and missile threats".[141] Accordingly, Defend (to counter the enemy's Engage step) may well need to include a capability to *return fire* against any person or object attacking or threatening to attack the LAWS. This is likely to be controversial, as such defensive attacks will not have benefitted from any human-led targeting process, except being considered in the abstract, as a possible risk arising from the operational environment; thus, they will be almost completely dependent on sensory hardware and control software.[142]

In that connexion, and in view of the absence of any direct mortal threat, it is important that deploying forces determine and clarify in advance the exact parameters for 'platform-defence'. On the one hand, Arkin's 'conservative use of lethal force' concept, in which a LAWS will return fire only *after* being fired upon, may be too little, too late for an aerial drone.[143] Arguably, a defensive but 'anticipatory' capability is needed. The inherent right of *individual* self-defence accorded to human soldiers cannot be applicable to LAWS, as this would be a category error. Moreover, the

---

[138] JP 3-0, III-35.

[139] Ibid., III-38.

[140] Amitai Etzioni and Oren Etzioni, 'Pros and Cons of Autonomous Weapons Systems', *Military Review* (May-June 2017), 73.

[141] JP 3-0, III-38.

[142] As noted in 5.2.5.3, this raises the possibility of fratricide, unintended escalation of a crisis, and civilian casualties due to adversaries manipulating defensive fires before hiding behind human shields.

[143] Arkin (n 93), 29, 46. This kind of "self-sacrifice to reveal the presence of a combatant" is arguably more apt for ground robots interacting with humans in a cluttered environment.

notions of 'hostile act' or 'hostile intent' arguably provide no help: as Gaston demonstrates, these ROE concepts – while conceptually reactive and threat-based – are in practice relatively time-distant.[144] Thus, they are often triggered by non-imminent threats, which may include threats to the broader *operation*, rather than the specific LAWS unit or its immediate strike mission.[145] Even more problematic is how ambiguous 'hostile intent' and 'hostile act' can be to apply, with the result that programming these concepts into nearer-term LAWS will make the latter prone to targeting errors.[146] Thus, an appropriate middle-ground will have to be sought, and it is for weapons designers to determine this based on the prevailing technology. At the very least, Electronic Warfare Capabilities may be installed and optimised for a given LAWS unit.[147]

5.3.3 Preliminary Conclusion on the Joint Targeting Cycles

As the above clearly demonstrates, both targeting cycles adopt a *methodical, step-by-step approach*, to maximise the chance of striking only legitimate targets, and to eliminate or minimise collateral effects. Moreover, *lawyers are an integral part of the process* to ensure compliance with LOAC norms and mission ROE.[148] As former General Counsel to the US Department of Defense Jennifer O'Connor recently pointed out, even when commanders need to make "rapid-fire decisions" in a dynamic cycle, a military lawyer will always be nearby to advise on issues of legality.[149] There are also major differences between the deliberate and dynamic cycles, which are most clearly seen in the *time available* for each one. To be sure, it works the other way around: the amount of time available determines which targeting cycle is to be used. As O'Connor puts it:

---

[144] See EL. Gaston, 'When Looks Could Kill: Emerging State Practice on Self-Defense and Hostile Intent', *Global Public Policy Institute Research Paper* (22 June 2017) <http://www.gppi.net/fileadmin/user_upload/media/pub/2017/gaston_2017_hostile-intent_web.pdf> accessed 7 July 2018.

[145] Ibid. Such broader threats are arguably best left to human deliberative thinking/controlled processing.

[146] Ibid., 11, 13, 16 and 20-21 (giving the examples of talking into a phone, shining a light at attacking forces, moving in range of combat assets, or even running away (hostile intent); or, digging a hole in the ground, driving through a perimeter, or speeding towards attacking forces (hostile act)).

[147] See Tony Gillespie and Robin West, 'Requirements for Autonomous Unmanned Air Systems Set by Legal Issues' (2010) 4 The International C2 Journal 23 (discussing Electronic Surveillance Measure (ESM) sensors and Electronic Counter Measure (ECM) jammers, to enable active and passive defences, respectively).

[148] Jennifer M. O'Connor, 'Applying the Law of Targeting to the Modern Battlefield', *Speech Delivered by DoD General Counsel to New York University School of Law* (28 November 2016) <https://www.defense.gov/Portals/1/Documents/pubs/Applying-the-Law-of-Targeting-to-the-Modern-Battlefield.pdf> accessed 7 July 2018.

[149] Ibid., 7-8.

> When there is time to plan in advance for a particular target, we [use] deliberate targeting. When we are reacting to an immediate need or attacking a target that is on the move to cause harm, we [use] dynamic targeting.[150]

Accordingly, another difference between the two targeting cycles that is implicit in O'Connor's comments is the *type of target* being pursued: whether it is a planned target (prosecuted via the deliberate cycle), or a target of opportunity (dynamic cycle).

For a LAWS deployment, however, both of these count as one from a total of three categories:

(1) **Specific and unique targets** (either *planned* or of *opportunity*) that have been individually prioritised for engagement through one of the targeting cycles.

(2) **Targets belonging to a broader set**, where the set has been approved through one of the targeting cycles, and will be executed through an autonomised (and compressed) dynamic cycle.

(3) **Persons or objects attacking or threatening to attack the LAWS**, which in turn drive the application of kinetic force in platform-defence, to preserve combat capability.

5.4 The Central Decision in Targeting

Roorda discusses the 'central decision' which, as briefly introduced above, occurs immediately before the moment where humans can no longer influence the direct violent effects.[151] This point of 'no return' varies from one weapon system to the next,[152] and may be effectuated by *commission* or by *omission*.[153] For a LAWS operating with a man *out*-of-the-loop, this will usually occur by commission at the point of launching the weapon system. For a system with 'remote check-in' it may occur by omission, where the weapons operator (WO) is satisfied with all engagements and allows the system to continue.[154] For multiple LAWS overseen by a single WO

[150] Ibid., 3.

[151] Namely, during *force execution* in the deliberate cycle, or the *engage* stage of the F2T2EA (dynamic) cycle.

[152] For a sniper rifle, the 'point of no-return' is the pulling of the trigger. For a fire-and-forget missile, it is the pressing of the launch button. For a cruise missile that is 'reprogrammable in flight', it is not the launch point, but the point at which reprogramming is no longer feasible.

[153] The sniper rifle and the fire-and-forget missile reach this point with positive activation; the reprogrammable cruise missile reaches it with no further action on the part of the WO, after activation.

[154] See 7.3.6.6.

who remains *on*-the-loop, the point of 'no return' is also passed by omission when the system is in flight, has 'positively identified' a target and is not manually overridden.[155] In all cases, the person making that final judgment must consider *whether the expected attack will comply with operational and legal requirements*.[156] His knowledge of the following five factors will influence this:[157]

(1) The situation at the time of the decision.
(2) The expected functioning and effects of the LAWS (including the WO's ability to provide the system with appropriate parameters).
(3) The possible changes in the situation between the decision and the kinetic effects.
(4) The accuracy of the intelligence used for these assessments.
(5) The operational and legal requirements.

For a decision to proceed with the attack, the combination of these factors must lead to the conclusion that the expected effects will remain within operational and legal boundaries.[158] Importantly, this also implicates the vast array of preparatory work and decisions made by JFCs and their battle staffs, as overseen and reviewed by JTCBs, during the first five phases of the deliberate targeting cycle. Should the WO/decision-maker at that point of 'no return' *not* be convinced that the expected effects will be operationally and legally compliant, he must either postpone the attack and re-plan it, or cancel it altogether.[159]

In the event that re-planning occurs, Roorda points to 13 specific options, of which the following are the most relevant to those deploying a LAWS:[160]

[155] Though it is acknowledged that the likely speed and responsiveness with which a LAWS will prosecute targets may effectively negate any meaningful decision by omission on the part of the WO. If so, then the true point of no return will be blurred, even if it still occurs mid-flight.

[156] Roorda (n 38), 161 (arguing that the moment at which control is relinquished over the direct outcome – and the human decision that allows the process to surpass this point – should be scrutinised more closely).

[157] Ibid., 161-62.

[158] Ibid., 162.

[159] Ibid.

[160] Ibid.

- Limiting operations that lead to dynamic targeting.
- Decreasing the spatial and temporal boundaries of operations.
- Increasing intelligence collection and imposing measures that provide clarity on the accuracy of the intelligence.
- Limiting operations to uncluttered environments, in which intelligence on the situation is more easily collected.
- Imposing measures that prevent changes to the current situation.
- Giving advance warning to the civilian population
- Testing and, if necessary, reprogramming the assigned LAWS.
- Operating the LAWS in remotely-piloted mode, or opting for other means.

Arguably, the combined effect of the detailed targeting cycles and the 'central decision', with its 'final judgment' factors and re-planning options, is to marshal a strong element of human judgment and control in US/NATO targeting, irrespective of the weapon system used.

5.5 Is There Meaningful Human Control Over Attacks Planned Under the Joint Targeting Cycles?

Recall that a LAWS will select and engage specific targets in Phase 5b (force execution) of the deliberate cycle, or through the entirety of the (compressed) dynamic cycle. The first of these two critical functions (select) will vary in nature depending on whether the mission is exclusively a targeted strike, or one that aims to attack targets belonging to a broader set. The former involves *unique* target parameters, hence little or no machine discretion to 'select'; the latter entails relatively broader target parameters with correspondingly wider machine discretion in 'discharging' the JFC's/WO's distinction obligation. In both cases, the machine is unsupervised at the point of kinetic attack, and this has raised concerns that human judgment and control are inherently lacking,[161] despite being a structural requirement of the law.

[161] Most prominently, Human Rights Watch, *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch, 2012), 7 (arguing that "[i]f this trend [of growing weapons autonomy] continues, humans could start to fade out of the decision-making loop, retaining a limited oversight role – or perhaps no role at all"). See also Markus Wagner, 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems' (2014) 47 Vanderbilt Journal of Transnational Law 1371, 1405 (arguing that "the use of AWS creates a number of complications precisely because individuals are removed from the decision-making process"). Others who oppose the removal of narrow-loop human judgment include those cited in Chapter 1, nn 21-23 therein.

5.5.1 A Strongly Human Element Within the Joint Targeting Cycles

Yet, as 5.3 reveals, this point of kinetic attack is preceded by a strongly deliberative process, the hallmark of which is extensive human-led planning.[162] The entire process is guided by in-depth technical expertise, combat experience and legal advice,[163] which becomes progressively granular as we move through the targeting cycle. This continues until the 'central decision' point, by which time the human WO will need to be satisfied that the launch of (or permission to continue with) an attack will be legally and operationally compliant; or else he must either postpone and re-plan, or cancel the attack.

Accordingly, while LAWS will be narrowly autonomous during the *force execution* stage, they will not be truly autonomous in the overall targeting process, which may be regarded as the 'wider loop' of human control.[164] It is during this process, and within the commensurate wider loop of decision-making, that meaningful and deliberative human control is exercised by the JFC and his battle staffs (and overseen by the JTCB). As mentioned in 4.5.5, they achieve this by a) imposing operational constraints, and b) setting conditions for their own judgment. This leaves the weapon system to pursue narrowly-defined targets with all the advantages and efficiency of automatic processing. Significantly, even some proponents of a LAWS ban share this view – at least in relation to targeted strikes – and have framed their approach to LAWS regulation to require weapon systems to be:

> [D]esigned to only attack particular [i.e. specific] acquired targets or specified locations known to the commander authorizing and the operator aiming and triggering the weapon.[165]

The corollary is again that LAWS *per se* will not 'comply with IHL/LOAC', in that the system itself will not necessarily (and will certainly not need to) distinguish

---

[162] Roorda (n 38), 163 ; Merel Ekelhof, 'Human Control in the Targeting Process' in Robin Geiß (ed.), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016), 73.

[163] O'Connor (n 148).

[164] Ekelhof (n 162), 73; Roorda (n 38), 163.

[165] Mark A. Gubrud, 'The Ottawa Definition of Landmines as a Start to Defining LAWS', *1.0 Human: Mark Gubrud's Weblog* (April 2018), 4 <http://gubrud.net/wp-content/uploads/2018/04/Landmines_and_LAWS.pdf> accessed 7 July 2018 (expressly stating, at 5, that MHC underpins the author's approach, by requiring that specific targets are already determined at the trigger ('central decision') point).

civilians from military objectives, nor to undertake proportionality assessments or any precautions in attack. Arguably, there is no legal requirement for a weapon system to do this.[166] LOAC merely mandates compliance with the above principles and is silent as to the *mode* of compliance, which can occur via system capabilities and/or the manner of employment and use.[167] A distinct but related point is that there is no legal distinction between effecting control of a weapon system through physical manipulation (e.g. soldier manually pointing a rifle at a target) or through a computer program (e.g. programmer setting parameters on when, where and what the system can engage).[168] So long as the required span of control can be expressed through a software program, and the system will operate with a reasonable degree of certainty in a given environment, it is potentially legally and operationally compliant.[169] Accordingly, it is possible for a system with relatively basic distinction capabilities to be deployed in a sparse battlefield, to engage a column of tanks or a bridge.[170] In such a scenario, the machine will certainly not 'select' tanks based on any deliberative understanding of the threat posed by such military objects, or of the operational value to be gained from destroying them. Instead, it will target a particular heat and shape profile, along with any other relevant signatures that are programmed into the system's control software, which the JFC and battle staffs (overseen by the JTCB) and the operator have determined will be unique to enemy tanks in the particular operational environment.

The real question must therefore be: "how well has the person deciding on the employment of a [LAWS] assessed the implications of its use?"[171] This again goes back to the targeting cycle and the central decision, and demands that those who plan and execute an attack are:

---

[166] In contrast with the views and analysis put forward by Human Rights Watch (n 161); and Noel E. Sharkey, 'The Evitability of Autonomous Robot Warfare' (2012) 94 International Review of the Red Cross 787.

[167] See, for example, Michael N. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) Harvard National Security Journal Features 1, 12; Jeffrey S. Thurnher, 'Examining Autonomous Weapon Systems from a Law of Armed Conflict Perspective' in Hitoshi Nasu and Robert McLaughlin (eds.), *New Technologies and the Law of Armed Conflict* (TMC Asser Press, 2014), 219.

[168] Lieutenant Colonel Christopher M. Ford, 'Autonomous Weapons and International Law' (2017) 69 South Carolina Law Review 413, 454.

[169] Ibid. Indeed, as was argued in 4.2.2, technical manipulation may practically enhance human control.

[170] Ibid.; Schmitt (n 167); Thurnher (n 167). See also 6.5.5.

[171] Roorda (n 38), 164.

(1) Fully informed on both legal and operational constraints.

(2) Familiar with the technical characteristics of the LAWS, as well as the complexity of the tasks being delegated to it.[172]

(3) Aware of the (civilian) risks presented by that system in the particular operational environment.[173]

(4) Have implemented all feasible precautionary measures to mitigate those risks.[174]

Consequently, the term 'critical functions' may be reconceptualised from the narrow *technical* process of target recognition, to the more deliberative *human-led* process of Joint Targeting, with its multiple stages and checks and balances.

Figures 5.5-5.8: Various targeting activities undertaken by a range of battle staffs, for US Special Operations Command. At all stages, metacognitive thinking, human judgment and due diligence are key. Source: Joint Chiefs of Staff, *Joint Publication 3-05.2*, 21 May 2003, I-1, I-4, II-12 and III-1.

5.5.2 A Note on the 'Individual Attack' Limitation

Recall from 4.3 that operational parameters – time and space of operation, quantity and complexity of tasks/engagements – must be set tightly enough to ensure MHC over 'individual attacks'. While the application of this concept is highly contextual

[172] Recall from 2.2.3.2 that task complexity depends on precision, tangibility, dimensionality and interaction.

[173] See 2.2.3.2 on environmental complexity.

[174] For example, pursuant to Article 57, AP I, and the customary norms restated in CIHL, Rules 15-21, or the more detailed list put forward by Roorda at (note and text accompanying) n 160.

and will depend on a number of other factors,[175] Roff and Moyes argue that it requires human judgment and control to be exercised on the *tactical* level of warfare, not just the strategic and operational levels:

> Broadening the concept of an 'attack' [beyond the tactical level] risks diluting the information available to human commanders as a basis for their legal and operational judgments to the point where their ability to predict outcomes becomes either non-existent or minimal.[176]

Accordingly, the authors consider that an 'attack' is a single "unit of analysis" in which human judgment and control is applied.[177] Breaching this will not only undermine the structure of LOAC, but will also mean LAWS deployments fail to take into account situational changes between weapons launch (the 'central decision') and kinetic effects.[178] Namely, such expansive deployments will likely fail the third of Roorda's five 'final judgment' criteria.[179] Moreover, they may lead to a situation where large operations are determined lawful on the basis of broad strategic/anticipated outcomes, while containing multiple actions that are individually unlawful.[180] This would be contrary to the spirit and purpose of Phases 1 and 5 of the deliberate targeting cycle, where broader goals are handed down and operationalised. For both of these reasons, the idea of an 'individual attack' does not conceptually add anything to current US and NATO targeting practice *in relation to targeted strikes*; so long as these are launched while the contextual information they rely upon remains relevant, to avoid unanticipated situational changes.[181] However, it does arguably reinforce these practices in the face of possible pressures to dilute the law, as autonomous technologies and loitering capabilities develop.[182]

---

[175] For example, it may also depend on the complexity of the battlefield, the sophistication of the weapon system, and the transparency or opacity of that system. See Heather Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons', *Briefing Paper for Delegates at the CCW Meeting of Experts on LAWS* (11-15 April 2016), 4 <http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf> accessed 7 July 2018.

[176] Ibid.

[177] Ibid.

[178] Ibid.

[179] See criteria at (note and text accompanying) n 157.

[180] Article 36, 'Key Elements of Meaningful Human Control', *Background Paper to Comments Prepared by Richard Moyes for the CCW Meeting of Experts on LAWS* (11-15 April 2016), 3 <http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf> accessed 10 June 2018.

[181] Roff and Moyes (n 175), 3.

[182] Ibid., 5. In this regard, see also 5.5.3.

Importantly, an 'individual attack' may still involve multiple acts of violence against multiple specific targets, so long as these are within sufficiently tight spatio-temporal bounds as to ensure predictability of outcomes.[183] Not only does this permit several targeted strikes in a single deployment, it also means the individual attack limitation is potentially consistent with tactical-level combat against broader target sets. So long as the JFC and the WO are reasonably confident that the contextual information being relied upon remains valid, and there is timely human action and a capacity for intervention, such that every engagement reasonably accords with their attack intentions. The parameters to ensure this are likely to vary from case to case, and it is arguably best left to the judgment of the commander who is closest to a given LAWS deployment.[184] Potential 'control systems'[185] to ensure autonomy does not go beyond an individual attack are discussed at 7.2.3 and 7.3.6. In addition to target parametric and spatio-temporal limits, they include upper engagement limits, remote dial-in, deployment restrictions and an open-ended category of 'workarounds'.[186]

Accordingly, the 'individual attack' limitation is potentially satisfied in both types of engagement, but is more likely to be an issue with tactical-level combat because of the focus on broader target sets. By definition, these lead to multiple attacks on numerous specific targets that conform to generalised parameters; often, the exact time and location of specific attacks are not known when planning. Thus, tactical-level combat may need to be controlled with a more conscious 'individual attack' limitation, to pre-empt the risk that it continues after situational changes have occurred. By contrast, a targeted strike involves extensive planning focused on a specific, unique and (often) fixed target, which is likely to count as an 'individual attack', subject to temporal restrictions to avoid unanticipated situational changes.

---

[183] Ibid. Current systems that fit this description include the sensor-fused weapon and the multiple-launch rocket system.

[184] Assuming the commander acts in good faith, which is a safe assumption in the case of US and NATO forces, as these often incorporate civilian risk mitigation in their strategic, operational and tactical mission goals.

[185] Article 36 (n 180), 3 (pointing out that restricting autonomous operation to individual attacks demands "some form of control system", to ensure a) human legal judgment is applied in each attack and b) the capacity for that judgment to be acted upon in a timely manner).

[186] Some of these effectively create multiple new 'central decision' points while the LAWS is on deployment.

5.5.3 Is there a Risk of Autonomising the 'Wider Loop'?

So far, the analysis has focused on MHC in the 'wider' targeting process, with tactical or 'narrow' autonomy only during *force execution* and restricted to 'individual attacks'. However, rapid technological change may well see machines and software algorithms taking over the entire deliberate targeting cycle, from formulating strategic goals to deciding when and where to strike.[187] Accordingly, Horowitz briefly considers the possibility of an 'autonomous operational system', which would involve LAWS planning military operations in a way that "completely replaces the human staff systems that plan military operations today".[188] Roff offers a more in-depth and speculative account of autonomising the wider loop by way of the *Strategic Robot Problem* where numerous self-contained, independently-targeting LAWS fail to coordinate, and cause command and control failures.[189]

However, while Roff's analysis focuses on the dystopian possibilities of wider loop autonomy, Horowitz largely dismisses the idea as being akin to science fiction, at least for the foreseeable future. More importantly, he argues that militaries are unlikely to even pursue such technologies, since States want to maintain control over the use of force.

> As the decision-making distance between the person activating the [autonomous operational] system and the system selecting and engaging targets increases, it makes fair accountability harder and makes a responsibility gap more likely…what that [L]AWS would do might be highly uncertain, disconnecting the person from the actual selection/engagement of targets…*This is the situation where meaningful human control…would be most at risk.*[190]

To be sure, Roff and Horowitz are discussing two different types of systems, with the former presenting greater command and control challenges. Nonetheless, both types of system will almost certainly undermine MHC and, as Horowitz points outs, human control in the wider loop is coterminous with military utility. This fact alone, which is

---

[187] Ekelhof (n 162), 74.

[188] Michael C. Horowitz, 'Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems' (2016) 30 Temple International & Comparative Law Journal 85, 96.

[189] Heather M. Roff, 'The Strategic Robot Problem: Lethal Autonomous Weapons in War' (2014) 13 Journal of Military Ethics 211.

[190] Horowitz (n 188), 96 (emphasis added).

supported by other authors with military experience,[191] will very likely militate against the fielding of any such systems.

Yet, none of this detracts from the possibility of a machine-dominated targeting process that *gradually* supplants human control in the wider loop.[192] Such 'creeping autonomy' is more likely to replace individual tasks that are currently undertaken by humans in Phases 1 to 5a of the deliberate cycle, with possible *automation bias* eroding MHC.[193] As Ekelhof argues, the possibility of autonomy with complex reasoning and machine learning *should* be considered in light of the broader targeting process, to assess how humans can remain in control in the event that such technologies are eventually adopted.[194]

In that regard, the US DoD's *Project Maven*[195] is a potential concern. This seeks, amongst other things, to refine AI and machine learning algorithms for computer vision, to detect and identify specific objects in 38 categories, especially in the fight against Islamic State.[196] More generally, *Maven* is tasked with automating the Processing, Exploitation and Dissemination (PED) of the ever-increasing volumes of full-motion video (FMV) drone feeds at Phase 2 of the deliberate targeting cycle, to reduce the burden on human analysts.[197] This is intended to increase actionable intelligence, and speed up decision-making through the rest of the targeting cycle, partly because analysts freed from routine object classification tasks are able to use their talents to perform higher-end analysis.[198] More generally, the greater speed, precision and accuracy of big data-processing stands to uncover relationships and

---

[191] For example, Ford (n 168), 451 ("As a matter of practice, militaries and commanders spend considerable time and money to maximize control over their weapon systems. Indeed, control is arguably the very essence of a military – whether control of troops, units, weapons, or munitions").

[192] Ekelhof (n 162), 74.

[193] See Chapter 3, (note and text accompanying) n 10 therein.

[194] Ekelhof (n 162), 74.

[195] See Deputy Secretary of Defense Memorandum, 'Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven)' (26 April 2017) <https://www.govexec.com/media/gbc/docs/pdfs_edit/establishment_of_the_awcft_project_maven.pdf> accessed 7 July 2018.

[196] Marcus Weisgerber, 'The Pentagon's New Algorithmic Warfare Cell Gets Its First Mission: Hunt ISIS', *Defense One* (14 May 2017) <http://www.defenseone.com/technology/2017/05/pentagons-new-algorithmic-warfare-cell-gets-its-first-mission-hunt-isis/137833/> accessed 7 July 2018.

[197] Ibid. (noting that up to 80% of analysts' time is normally spent on mundane administrative tasks, like staring at FMV feeds, manually labelling objects and entering data into spreadsheets; yet all of this can be automated).

[198] Weisgerber (n 196).

dependencies that humans would likely fail to recognise. Thus, *Maven* is not a weapon as such, but is a decision-support system for target development and enhanced situational awareness. Its rapid success is illustrated by the fact that one week into trials to identify people, vehicles and types of buildings, the accuracy of its algorithms improved from 60% to 80%.[199] Consequently, a move towards greater levels of autonomy in the wider loop may not be unrealistic, both in the US[200] and beyond.[201]

Clearly, if the project goals are realised as described, *Maven* will enhance human judgment and control in the wider loop. On the other hand, as was seen in Chapter 2, complex algorithms tend to behave like 'black boxes', with inscrutable and non-intuitive outputs.[202] Should this lead to sub-optimal or erroneous selections for human analysts to see, this will deteriorate situational awareness and may undermine human control within the wider targeting process.[203] Worse still, object classification remains inherently vulnerable to adversarial examples, which can fool systems into misclassifying objects.[204] Should this brittleness be exploited by adversaries to manipulate target development and to direct attacking forces towards protected persons and objects, this may weaken human control to catastrophic levels. Much will depend on how well such wider loop autonomy is tested, evaluated and integrated before formal rollout.[205]

---

[199] Marcus Weisgerber, 'The Pentagon's New Artificial Intelligence is Already Hunting Terrorists, *Defense One* (21 December 2017) <http://www.defenseone.com/technology/2017/12/pentagons-new-artificial-intelligence-already-hunting-terrorists/144742/> accessed 7 July 2018.

[200] Jack Corrigan, 'Three-Star General Wants AI in Every New Weapon System', *Defense One* (3 November 2017) <https://www.defenseone.com/technology/2017/11/three-star-general-wants-artificial-intelligence-every-new-weapon-system/142239/?oref=defenseone_today_nl> accessed 7 July 2018 (citing Lt. Gen. Jack Shanahan, who is in charge of the AWCFT, describing *Maven* as the "pathfinder" that will "spread AI techniques to the rest of the DoD").

[201] See 'Intelligence Technology to Keep Joint Force Command One Step Ahead of Adversaries', *Ministry of Defence News* (17 July 2018) <https://www.gov.uk/government/news/intelligence-technology-to-keep-joint-force-command-one-step-ahead-of-adversaries> Accessed 18 July 2018 (reporting on Phase 2 development of the 'predictive cognitive control system' for the UK's Joint Force Command. The project goal is to "take[…] a broad range of incredibly complex data, beyond the ability of analysts to simultaneously comprehend, and through the use of Deep Learning based neural networks…make confidence-based predictions of future events and outcomes of direct operational relevance to Defence Users").

[202] See 2.5.3.3 and 2.5.6.

[203] See Dustin Lewis, Naz Modirzadeh and Gabriella Blum, 'The Pentagon's New Algorithmic-Warfare Team', *Lawfare* (26 June 2017) <https://www.lawfareblog.com/pentagons-new-algorithmic-warfare-team> accessed 7 July 2018 ("Of particular concern are technologies whose 'choices' may be difficult – or even impossible – for humans to anticipate or unpack or whose 'decisions' are seen as 'replacing' human judgment").

[204] See 2.5.6.

[205] Merel AC. Ekelhof, 'Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting' (2018) 71 Naval War College Review 61, 82-85.

5.5.4 MHC in Weapons Design and Development

As was apparent in Chapter 4, the extent of MHC during downstream/*in bello* targeting processes is necessarily influenced by the design, development and testing of weapon systems upstream. Such *ante bellum* decisions can facilitate MHC in the targeting process by making systems more *predictable*, *reliable* and *transparent*, thereby helping to mediate the commander's/WO's intentions and actions.[206] This was discussed at 4.5, in relation to the elements of MHC. Moreover, as discussed in 4.2, system design may *effectuate* downstream actions,[207] or it may *prohibit* those actions in line with human judgment and control.[208]

Upstream design decisions may also entail technical and legal considerations about distinction proxies; for example, the robustness of image matches, the number and quality of signatures that must be reconciled with pre-programmed parameters,[209] and the statistical 'confidence thresholds' that must be satisfied for weapons release.[210] All these decisions heavily influence the degree of MHC that can be properly exercised during the targeting process.

Accordingly, MHC undoubtedly plays a role in upstream processes. Yet, there is no common agreement at present on the form of human control that designers and developers should embed into systems. As argued in 4.4.3, however, determining the 'core' MHC obligations of WOs – a much easier starting point – may generate fruitful leads for designers and developers. In turn, this would bring other, more formal legal decision-making processes into play; most notably, those involved in the legal review mechanism under Article 36, AP I.[211]

---

[206] Noorman (n 103), 812.

[207] Via user interface designs and data connectivity (4.2.1), or automated processes that execute human instructions relatively more accurately and precisely (4.2.2).

[208] Via software suppressors that automatically veto unsafe or unlawful human action (4.2.3).

[209] See 2.5.4.1 on standard ATR approaches.

[210] Alan Backstrom and Ian Henderson, 'New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews' (2012) 94 International Review of the Red Cross 483, 495 and 508-512.

[211] Ekelhof (n 162), 75; Article 36, 'Autonomous Weapon Systems : Evaluating the Capacity for 'Meaningful Human Control' in Weapon Review Processes', *Discussion Paper for the Group of Governmental Experts Meeting on LAWS* (13-17 November 2017) <http://www.article36.org/wp-content/uploads/2013/06/Evaluating-human-control-1.pdf> accessed 7 July 2018. See also 6.3.

5.6 Conclusion

This chapter has examined targeting practices under US and NATO doctrine, and has linked these to three distinct categories of target and engagement:

(1) **Specific and unique targets** that have been individually authorised for engagement, and which drive *targeted strikes*.

(2) **Targets belonging to a broader set**, which drive *tactical-level combat* that does not need engagement authority for each specific target.

(3) **Persons or objects attacking or threatening to attack the LAWS**, which in turn drive the application of kinetic force in *platform-defence*, to preserve combat capability.

Going forward into Chapters 6 and 7, the first two of these crucial distinctions will inform the application of specific targeting rules to the deployment and use of LAWS. As will be seen there, both categories are potentially legally compliant but targeted strikes will clearly afford the greatest level of MHC; thus, they present the greater likelihood of LOAC compliance, *ceteris paribus*.

As a structural requirement of the law, MHC plays a strong role in both the (upstream) design, development and legal review processes, and the (downstream) Joint Targeting Cycles. With the latter encompassing numerous sequential steps to develop, nominate and prioritise targets, and further steps for mitigating civilian risk, there are clear operational and legal safeguards to ensure that human judgment and control is not displaced in the wider loop. This permits narrow loop autonomy during Phase 5b (force execution), subject to the 'individual attack' limitation, which is also a structural feature of the law, and cannot be abrogated. Concretely, the Joint Targeting Cycles will effectuate MHC in a LAWS context by setting operational parameters and the conditions of judgment, to ensure the technology that has been approved as predictable, reliable and transparent in various upstream processes, is put to a lawful use downstream.

Yet, two problems remain. First, as noted above, autonomy will not be limited to Phase 5b, but is becoming a more pervasive influence throughout the targeting process. This includes Phase 2, which raises questions about the nature and extent of human judgment during the crucial target development process. Namely, whether such

judgments are being enhanced and utilised in the most productive manner; or truncated and supplanted by potentially erroneous algorithmic selections. Much will depend on how effectively projects such as *Maven* are tested, evaluated and implemented to optimise the human-machine team. Second, the concept of 'individual attack' is potentially challenging: it may constitute multiple acts of violence against multiple specific targets, and its application is likely to be context-specific. Thus, it does not offer any bright-lines, and it may well give rise to a penumbra of uncertainty. As noted above, autonomous and loitering technologies are likely to develop and push the notion of an 'attack' towards being conceptualised at broader levels. It is therefore important that commanders are guided by 'predictability of outcome' in the circumstances, and not 'technological viability', when determining an individual attack. Of course, this challenge will be relatively easier to meet with targeted strikes, save for any remaining technological pressure to expand the spatio-temporal scope of operation.

Chapter 6
Targeting Law I: Can LAWS Be Deployed in Compliance with the Principle of Distinction?

6.1 Introduction

The previous four chapters have developed the following argument.

- Lethal autonomous weapon systems (LAWS) will follow a set of technical processes, which may operate with super-human accuracy and precision, or with brittleness and potential failure, depending on context and circumstances (Chapters 2 and 3).
- In contradistinction, international humanitarian law (IHL)/law of armed conflict (LOAC) presupposes sentience and self-awareness, for imposing legal obligations; and human (autonoetic) metacognition, for its effective application in armed conflict. LAWS will possess none of these characteristics in the near-term. (Chapter 3).
- Accordingly, LAWS may only be lawfully deployed with a contextual application of meaningful human control (MHC),[1] to ensure structural and substantive compliance with the LOAC targeting rules (Chapter 4).
- The targeting process – at least in a US/NATO context – affords substantial opportunity to ensure MHC, because of its highly deliberative planning processes; so long as autonomy does not supplant human judgment in the wider loop, and the 'individual attack' limitation is maintained over and above mere technical viability (Chapter 5).

In addition, 'narrow loop' autonomy will *reassign* some targeting decisions between human actors, requiring them to be made relatively *earlier* and at *locations further removed* from the intended strike site. All things being equal, this will weaken the causal nexus between human decisions and specific battlefield outcomes (Chapter 3). Again, the US/NATO targeting process addresses this, as it incorporates substantial precautionary steps in order to minimise civilian risk.

[1] Namely, MHC should be applied in a way that facilitates compliance with the existing LOAC norms. 'Human control' should not become the object of a LAWS deployment, lest it weakens compliance with the LOAC rules and principles. See 4.5.6.

With these in mind, the following two chapters will apply the LOAC targeting rules to LAWS deployments, in a US/NATO context. After an introduction to the normative LOAC framework in 6.2, there is a brief note on weapons law in 6.3 and an overview of targeting law at the high level in 6.4. Subsequently, the chapter moves onto a more detailed application of the principle of distinction, both generally (6.5.1) and separately to persons (6.5.2) and objects (6.5.3). Throughout, it will be seen that legal compliance is primarily a concern for commanders and weapons operators (WOs), not the machine itself. This is underscored in 6.5.4, which addresses a crucial presumption in favour of civilian protection: the rule of 'doubt'. Namely, the 'fog of war', which gives rise to incomplete or inconclusive information, will sometimes demand metacognitive thinking to recognise *when* there is *enough* doubt to hold fire. Hence, outside relatively concrete situations, the requisite level of doubt to presume civilian status may have to be assessed by a human somewhere in the loop. Finally, 6.5.5 provides some closing thoughts on ensuring that LAWS deployments will comply with the principle of distinction, mainly through human-machine teaming and a 'division of labour'.

6.2 The Normative IHL/LOAC Framework

IHL/LOAC largely regulates the conduct of hostilities on the battlefield.[2] According to Sassóli, Bouvier and Quintin,[3] its principal aim is to limit human suffering during armed conflict by:

- protecting persons who are *not*, or who are *no longer*, directly participating in hostilities; and
- restricting the means and methods of warfare to those which are *necessary* to achieve the *legitimate* aim of the conflict, namely, to *weaken the military potential of the enemy*.

From this, we see a key overarching theme of LOAC, which was discussed in Chapter 4: the 'compromise' between military necessity and humanity (MN-H).[4] The former permits all lawful measures intended to engage and defeat the enemy as quickly and

[2] This is in contrast to *jus ad bellum*, which seeks to regulate the resort to war.

[3] Marco Sassóli, Antoine A. Bouvier and Anne Quintin, *How Does the Law Protect in War?: Cases, Documents and Teaching Materials on Contemporary Practice in International Humanitarian Law, Vol. I* (3rd ed., ICRC, 2011), 1.

[4] Michael N. Schmitt, 'Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance' (2010) 50 Virginia Journal of International Law 795.

as efficiently as possible.[5] The latter forbids the infliction of any *further* suffering, injury or destruction that is *not* necessary to accomplish a legitimate military purpose;[6] thus, humanity may be seen as the 'logical inverse' of military necessity.[7] Furthermore, the MN-H balance represents a categorical rejection of the 19th Century German doctrine of *Kriegsraison*,[8] which held that necessity in war overrules the manner of warfare.[9] Yet, the net result of combining the two principles is not necessarily a compromise, as each one may complement the other to reinforce military effectiveness;[10] this is certainly true in some areas of LOAC,[11] though not in all.[12] As discussed in 4.3.3, neither of these principles exists as a separate LOAC norm, but rather they pervade and undergird all other rules that regulate the conduct of hostilities.[13] These rules are mostly contained in Additional Protocol I[14] (AP I) and in the corresponding rules of customary law,[15] a large proportion of which converge around the principles of distinction and proportionality. The current chapter will focus on the former principle, while Chapter 7 will address the latter.

6.3 A Brief Note on Weapons Law

Any complete legal analysis of a weapon system under LOAC will comprise two distinct strands: whether the weapon *itself* is lawful (weapons law); and whether its

---

[5] US Department of Defense, *Law of War Manual* (DoD, 2015; December 2016 Update) (hereafter, *US DoD Manual*), §2.2; UK Ministry of Defence, *The Manual of the Law of Armed Conflict* (OUP, 2004) (hereafter, *UK MoD Manual*), §2.2; *United States* v. *List (Wilhelm) et al. (The Hostage Case)* Case No. 7, 19 February 1948 (1950) 11 TWC 1230, 1253-56.

[6] Yoram Dinstein, *The Conduct of Hostilities Under the Law of International Armed Conflict* (3rd ed., CUP, 2016), 8-12; Schmitt (n 4). See also *UK MoD Manual*, §2.4; *US DoD Manual*, §2.3.1.

[7] *US DoD Manual*, §2.3.1.1.

[8] *The Hostage Case*, 1256.

[9] Geoffrey Best, *Humanity in Warfare: The Modern History of the International Law of Armed Conflicts* (Methuen, 1983), 172-79 (discussing the controversial rise of the doctrine in the writings of Carl Lueder).

[10] For example, sparing the incapacitated enemy combatant affords him a chance to recover outside of hostilities, and it conserves ammunition for attacking more active combatants.

[11] Yves Sandoz, Christophe Swinarski and Bruno Zimmermann (eds.), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Convention of 12th August 1949* (Martinus Nijhoff, 1987) (hereafter, *AP I Commentary*), ¶ 1958 (noting that in relation to the principle of distinction and the targeting accuracy of long-range missiles, "military interests and humanitarian requirements coincide" to ensure that the "margin of error is gradually reduced").

[12] For example, the principle of proportionality is, by definition, a case of compromise. See 7.2.

[13] Schmitt (n 4).

[14] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3.

[15] Restated in Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law, Vol. 1: Rules* (CUP, 2005) (hereafter, CIHL). All Rules available at: <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul> accessed 10 June 2018.

actual *use* is lawful (targeting law).[16] The focus of this thesis is on targeting law and practice, by which point US/NATO commanders can safely assume all weapons in their arsenal are lawful, subject to prescribed restrictions on use. Weapons law will generally not pose a significant barrier to the adoption of LAWS, though it is worth briefly addressing this area for the sake of completeness.

Article 35(2), AP I, prohibits weapons that are "of a nature to cause superfluous injury or unnecessary suffering"; Article 51(4)(c) prohibits those the effects of which cannot be limited, and are likely to spread from combatants to civilians without distinction; and Article 35(3) prohibits weapons that "are intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment".[17] Crucially, these norms relate to the *effects* of the *munition* or *projectile* that a weapon system fires or launches, not to its guidance system.[18] Namely, autonomy is a *manner of engagement* and cannot in itself be 'of a nature' to inflict the above unlawful effects.[19] Hence, the aforementioned rules are inapplicable to LAWS, unless the system is deployed to fire or launch the offending munitions.[20]

Article 51(4)(b), AP I, prohibits weapons that are 'inherently indiscriminate', in that they amount to a "means of combat which cannot be directed at a specific military objective".[21] This rule also concerns the "very nature or design" of a weapon, not its use in a given engagement.[22] Such weapons are unlawful *per se* as they will routinely strike combatants, civilians, military objectives and civilian objects without distinction.[23] The *Commentary* on the *Air and Missile Warfare Manual*[24] specifically

---

[16] *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion) [1996] ICJ Reports 226 (hereafter *Nuclear Weapons* AO). The two legal regimes roughly correspond with the upstream/downstream (or *ante bellum*/*in bello*) distinction introduced in Chapter 4.

[17] Articles 35(2)-(3) and 51(4)(c), AP I; CIHL, Rules 70, 44 and 12.

[18] William Boothby, 'Dehumanization: Is There a Legal Problem Under Article 36?' in Wolff Heintschel von Heinegg, Robert Frau and Tassilo Singer (eds.), *Dehumanization of Warfare: Legal Implications of New Weapons Technologies* (Springer, 2018), 38-39.

[19] Michael N. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) Harvard National Security Journal Features 1, 9.

[20] Ibid.

[21] Article 51(4)(b), AP I; CIHL, Rules 12 and 71

[22] Dinstein (n 6), 72.

[23] Schmitt (n 19), 10.

[24] Program on Humanitarian Policy & Conflict Research at Harvard University (HPCR), *Commentary on the HPCR Manual on International Law Applicable to Air and Missile Warfare* (v2.1, Harvard College, 2010) (hereafter, *AMW Manual Commentary*). See also the Manual to which this commentary

affirms that the principle of distinction (see 6.5) and its application to weapons design extends to autonomous systems.[25] Accordingly, Article 51(4)(b) cannot mean that 'autonomy' as a concept is inherently unlawful. So long as a *given* system can be supplied with sufficiently reliable and accurate data on target parameters as to enable it to pursue a specific military objective *in the operational environment for which it is intended*, it will not be indiscriminate by nature; thus not unlawful *per se*.[26]

Therefore, much depends on whether the guidance and target recognition systems in a *given* LAWS reach a legally acceptable standard of discrimination. Namely, the main weapons law issue is in the realms of testing and evaluation, and legal review under Article 36, AP I:[27] to determine when a given combination of (multi)sensed data indicates the presence of a military object; and to determine whether an *individual* system has reached the required degree of confidence, with an acceptable failure rate for a LAWS to positively identify such an object.[28] In this connexion, it should also be emphasised that *as a matter of law* autonomous systems *need not perform better* than manually-operated systems.[29] A risk of targeting error, which is at least common to man and machine, will not in itself render a LAWS unlawful.[30]

---

relates: HPCR, *HPCR Manual on International Law Applicable to Air and Missile Warfare* (Harvard College, 2009) (hereafter, *AMW Manual*).

[25] Ibid., Rule 17(a), ¶ 3 ("The sensors and computer programs [of a LAWS] must be able to distinguish between military objectives and civilian objects, as well as between civilians and combatants").

[26] Ibid., Rule 39, ¶ 4.

[27] Or under the corresponding rules of customary law. For an application of the legal review criteria to LAWS, see Boothby (n 18), 38-45.

[28] Alan Backstrom and Ian Henderson, 'New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews' (2012) 94 International Review of the Red Cross 483, 508-09.

[29] Schmitt (n 19), 12; Marco Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified' (2014) 90 International Law Studies 308, 320. See also *AMW Manual Commentary*, Rule 39, ¶ 4 ("In case of autonomous systems…The performance of the sensors and the program identifying lawful targets must be comparable to that of manned aircraft or to that of remotely-piloted (i.e. non-autonomous) UCAVs").

[30] While it is beyond the scope of this thesis, it is worth noting that problems under Article 36 are less to do with the legal review criteria itself, and more about the 'certification problem'. Namely, the near-infinite state possibilities of complex autonomous systems make it near-impossible to test every interaction that the systems might have in a complex operational environment, and therefore to fully evaluate lawfulness. This problem is compounded by the brittleness of AI and the 'black box' issue highlighted in Chapter 2. That said, there are currently efforts aimed at resolving the certification problem, for example, via 'life-cycle testing and evaluation'; and there is always the option that legal review panels may impose deployment restrictions and conditions of use (specific precautionary stipulations), where a particular LAWS system cannot be demonstrated to be lawful in a particular context. Thus, while acknowledging the practical difficulties brought up by certification and review, the current and next chapter will proceed on the assumption that such matters will ultimately be addressed.

6.4 Targeting Law: An Overview

In contrast to weapons law, targeting law determines whether the *actual use* of a weapon in battle and the actual *targeting process* is lawful. In that regard, Boothby defines 'targeting' as a broad process encompassing both *planning* and *execution*.[31] In a US/NATO context, this entails the detailed six-phase deliberate targeting cycle (for attacks planned more than 24 hours away); or the six expedited steps of the dynamic cycle (for attacks being planned within the current 24-hour cycle).[32]

As above, the law of targeting reflects the military necessity-humanity (MN-H) balance, most vividly in the principles of distinction (6.5) and proportionality (7.2). As will also be seen in 7.3, even the rules requiring precautions in attack – ostensibly stronger expressions of the principle of humanity – are structured to represent this delicate balance through the requirement of feasibility.

Hereon, the thesis will apply the LOAC targeting rules – both treaty-based[33] and customary[34] – to potential LAWS deployments, bearing in mind:

- the technical nature of autonomous weapons;[35]
- the obligation to exercise MHC over individual attacks;[36]
- the human-machine cognitive differences;[37]
- the Joint Targeting process,[38] which itself is an extensive form of MHC;[39] and
- the distinction between targeted strikes (TS) and tactical-level combat (TLC);[40] the latter involving relatively more machine discretion, while the former affords relatively more human control and human-specification of outcomes.

[31] William H. Boothby, *The Law of Targeting* (OUP, 2012), 4.
[32] See Chapter 5, especially 5.3.
[33] Principally, under AP I.
[34] As restated in the CIHL Rules.
[35] See Chapters 2 and 3.
[36] See Chapter 4.
[37] See 4.2.
[38] See 5.3.
[39] See 5.5.
[40] See 5.2.5.

6.5 The Principle of Distinction

The principle of distinction, "the most fundamental pillar of [LOAC]",[41] requires the Parties to an armed conflict to:

> …at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly [to] direct their operations *only* against military objectives.[42]

Thus, the principle imposes a bifurcated and cumulative obligation: to refrain from attacking civilians or their objects, and to ensure that all attacks are aimed at military objectives.[43] This clearly assumes and requires meaningful human control (MHC) over individual attacks.[44]

Distinction has been described by the International Court of Justice (ICJ) as an "intransgressible" principle of the LOAC.[45] While the Court did not elaborate on the meaning of this term, it is "more than probable" that it was intended to indicate that distinction is a norm of *jus cogens*.[46] That is, it may be a 'peremptory norm' of international law from which no derogation is ever permitted.[47] Not surprisingly, the principle also forms part of customary international law, thus it binds all States, whether or not they are Party to AP I.[48]

---

[41] Dinstein (n 6), 72 (adding that the principle reflects the history of LOAC as "a sustained effort to ensure that civilians…are protected from the havocs of war").

[42] Article 48, AP I (emphasis added). This is the 'Basic Rule', which is operationalised in subsequent Articles.

[43] *AMW Manual Commentary*, Rule 10(a), ¶ 3.

[44] See 4.3 on the structural legal argument.

[45] *Nuclear Weapons* AO, ¶ 79.

[46] Jean-Françoise Quéguiner, 'The Principle of Distinction: Beyond an Obligation of Customary International Humanitarian Law' in Howard M. Hensel (ed.), *The Legitimate Use of Military Force: The Just War Tradition and the Customary Law of Armed Conflict* (Routledge, 2016), 171 (though noting that this is far from explicit, and requires an element of "reading between the lines" to detect the Court's "implicit yet clear affirmation" of the *jus cogens* character of the principle of distinction).

[47] Article 53, Vienna Convention on the Law of Treaties (adopted 23 May 1969, entered into force 27 January 1980) 1155 UNTS, 331; Thomas Weatherall, *Jus Cogens: International Law and Social Contract* (CUP, 2015), 5-8. Independently of the *Nuclear Weapons* AO, Quéguiner, ibid., argues that three other reasons support the idea that distinction is a norm of *jus cogens*: a) the fact that Article 48 is entitled the 'Basic Rule', b) its location at the first Section of Part IV, a Chapter devoted to the Section's field of application, c) the fact that Article 48 requires distinction "at all times". Collectively, these demonstrate that the principle of distinction was intended to permeate all of the rules on the conduct of hostilities, and at all times.

[48] *Nuclear Weapons* AO, ¶ 79; Quéguiner, ibid., 164-67. See also the customary rules restated in CIHL, Rules 1 and 7. As a corollary, it may be argued that the requirement of MHC over individual attacks is also customary and it enjoys the status of *jus cogens*.

In targeting law, the broad principle enshrined in Article 48 is operationalised in a number of subsequent AP I rules. These apply to persons and objects, both separately and together in some general provisions.

6.5.1 General Provisions

On the face of it, the black-letter text of Article 51(4)(a), AP I, would appear to sound the death knell for autonomous lethal targeting, in all but TSs. This rule prohibits attacks "which are not directed at a *specific* military objective".[49] Hence, Dill argues that "[f]ailure to have a *specific* military objective *in mind* and to *direct one's weapon against it*…is…a violation of the principle of distinction";[50] other commentators have taken a similar stance.[51] Given that LOAC imposes obligations on humans and not on machines,[52] Article 51(4)(a) would therefore appear to make any LAWS deployment to engage in TLC unlawful, with wording that can arguably be seized upon by ban proponents.[53] However, on closer examination the prohibition has been interpreted differently to this. For example, Dinstein argues that the key to a finding of 'indiscriminate attack' under Article 51(4)(a) is the "nonchalant state of mind of the attacker".[54] Boothby illustrates this with the examples of "blind firing" a rifle, or "the firing of artillery *without making any attempt to direct the munition*".[55] Dinstein further adds bombing raids from high altitudes, in conditions of poor visibility and inclement weather, and utilising no precision guidance mechanism.[56] Indeed, the gist of Article 51(4)(a) seems to be more about the WO's diligence (or "the attacker's indifference"[57]) than the technical working of the weapon system.

---

[49] Article 51(4)(a), AP I (emphasis added).

[50] Janina Dill, *Legitimate Targets? Social Construction, International Law and US Bombing* (CUP, 2015), 75 (emphasis added).

[51] Remarks by Françoise J. Hampson, 'Proportionality and Necessity in the Gulf Conflict' (1992) 86 Proceedings of the Annual Meeting (American Society of International Law) 45, 49 (asserting there is a "requirement that *each potential target* be examined *separately*", thus "[i]t is *not possible* to have a *class* of targets") (emphasis added).

[52] See 3.2 and 4.3.1.

[53] See the approach to LAWS taken by Gubrud (a prominent ban proponent), in 5.5.1.

[54] Dinstein (n 6), 147.

[55] Boothby (n 31), 92 (emphasis added).

[56] Dinstein (n 6), 148.

[57] *AMW Manual Commentary*, Rule 13(a), ¶ 2.

Accordingly, in programming and deploying a LAWS with generalised but *appropriate* parameters, and with automatic target recognition (ATR) systems that are apt for the operational environment, commanders and their battle staffs will arguably not be violating Article 51(4)(a). If they intend, and can reasonably expect, that the system will *ultimately* be directed towards a specific and lawful target during TLC,[58] the question of human *versus* algorithmic determination of that target will be immaterial,[59] and the attack should not be construed to be indiscriminate. This is also underscored in the *AMW Manual Commentary* in relation to 'beyond-visual-range' attacks,[60] and it highlights the importance of a reasonable interpretation and application of the LOAC prohibitions, in accordance with the context at hand.[61]

Paragraph (5)(a) of the same Article prohibits the World War II practice of 'target area' bombings. This is defined as:

> [A]n attack by bombardment…which treats as a single military objective a number of *clearly separated and distinct* military objectives located in a[n]… area containing a…*concentration of civilians or civilian objects*.[62]

On the one hand, compliance with this rule may be a relatively simple matter of programming and deployment, which can be satisfied in the context of the deliberate targeting cycle. Namely, where it is possible for a manned weapon system to engage a number of targets individually (for example, in a series of discrete TSs), this should be

---

[58] The reasonableness of such expectation will partly depend on target parameters and spatio-temporal limits.

[59] Jeffrey S. Thurnher, 'Means and Methods of the Future: Autonomous Systems' in Paul AL. Ducheine, Michael N. Schmitt and Frans PB. Osinga (eds.), *Targeting: The Challenges of Modern Warfare* (TMC Asser Press, 2016), 186; *AMW Manual Commentary*, Rule 39, ¶ 4.

[60] *AMW Manual Commentary*, Rules 7(c), ¶ 3 and 13(b), ¶ 8 (pointing out that 'beyond-visual-range' weapon systems "are lawful", hence "not as such indiscriminate when their employment permits distinguishing military objectives and combatants from civilians and civilian objects…through sensors on the weapon itself, or through external guidance"). See also 4.5.6 on the likely customary status of precision-guided munitions.

[61] In this connexion, it should be noted that Dill's comment, (text accompanying n 50), was not focusing on LAWS, but was made in the context of the bifurcated obligation inherent in the principle of distinction. Seen in this context, the author would not necessarily object to the autonomous selection and engagement of a specific military target that a reasonable combatant would also have engaged in the case of manned targeting. By contrast, Hampson's comment (at n 51) was made in the context of the temporal aspect of the definition of military objective under Article 52(2), which may in some narrow circumstances pose problems for autonomous attack; on which, see 6.5.3.3. Nonetheless, the point remains that the black-letter text of the general LOAC provisions must be subject to reasonable interpretation and application to LAWS, in light of the underlying goal that the provision aims to achieve, or the mischief that it aims to eliminate.

[62] Article 51(5)(a), AP I (emphasis added); *AMW Manual Commentary*, Rule 13(c), ¶ 3; CIHL, Rule 13.

the default programming option for a LAWS, rather than deploying it to engage those targets *en masse*.[63] Conversely, Boothby has questioned the ability of ATR technologies to detect targets separately, especially in heavily built-up areas where tight comingling may require evaluative and metacognitive judgments, to which LAWS are not easily suited.[64] Concretely, where military targets are in fact separated, yet the machine-imperceptibleness of urban clutter causes the ATR to perceive them as a single unit, there may be an inadvertent violation of Article 51(5)(a). This is more likely to be a risk in TLC, which does not benefit from pre-strike deliberations on specific targets. But whether the risk does in fact materialise depends on the capabilities of the ATR compared with existing manned systems; the prevailing operational environment; and the robustness of the targeting process in matching one with the other.

In that regard, recall Figure 2.5 in 2.5.2, which showed a recurrent neural network successfully distinguishing neatly laid-out and clearly separated items on a breakfast table. Where ATR capabilities are limited to this, they may fail to adequately distinguish overlapping objects in a cluttered environment,[65] and may possibly violate Article 51(5)(a) if deployed. More recent research, however, has improved the accuracy of systems by training them to recognise full objects (the 'ground truth') and to estimate the full physical presence of those objects (as well as their separation from each other) more accurately, when partially concealed or obscured.[66] An example of the results is shown in Figure 6.1, below.

---

[63] Michael N. Schmitt, 'Autonomous Weapons Systems and International Law', *LENS Conference 2016: Autonomous Weapons in the Age of Hybrid War* (27 February 2016) <https://www.youtube.com/watch?v=b5mz7Y2FmU4> accessed 11 August 2018.

[64] The author argues that the "clearly separated and distinct" criterion requires the ATR of a machine to evaluate the relative positions of military objectives; and to assess the similarity or otherwise of their concentration with the concentration of civilians or civilian objects in the same 'locality'. This involves the comparison of dissimilar items – each one being of a different size, significance or value – thus, it "presuppose[s] the involvement of a human brain" with metacognitive functions. See William H. Boothby, *Conflict Law: The Influence of New Weapons Technology, Human Rights and Emerging Actors* (TMC Asser Press, 2014), 109-10.

[65] Tucker Davey, 'How AI Handles Uncertainty: An Interview with Brian Zeibart', *Future of Life Institute News* (15 March 2018) <https://futureoflife.org/2018/03/15/how-ai-handles-uncertainty-brian-ziebart/> accessed 13 May 2018.

[66] Sima Behpour, Kris M. Kitani and Brian D. Ziebart, 'ADA: A Game-Theoretic Perspective on Data Augmentation for Object Detection' (12 December 2017) <https://arxiv.org/pdf/1710.07735v2.pdf> accessed 13 May 2018.

Figure 6.1: A relatively complex environment in which the system attempts to classify the full presence (ground truth) of each person and object, despite some being obscured. Source: Davey (n 65).

This kind of training to separate each 'ground truth' may help to avoid inadvertent target area bombing. Nonetheless, it will be for commanders and their staffs in the targeting process to make the appropriate judgments on the technology and context when planning multiple TSs in close proximity, or when deploying a LAWS for TLC in a cluttered area.[67]

6.5.2 Persons

In line with the bifurcated nature of the distinction principle, LAWS will need to categorise persons either as combatants, who may be directly attacked; or as civilians, or other protected persons, who must be spared and protected from direct attack.

6.5.2.1 Active Combatants

6.5.2.1.1 The General Position

Combatants are generally members of the conventional armed forces who have the right to participate directly in hostilities,[68] or members of "other militias and…volunteer corps" who meet certain conditions,[69] and who will be engaged during TLC. These persons may be detectable to ATR via their distinctive uniform and

[67] Of course, for such systems to even be available for deployment will first require their technical certification and legal review, as per n 30.

[68] Article 43(2), AP I; *AMW Manual Commentary*, Rule 10(b)(i), ¶ 2; CIHL, Rule 3.

[69] Article 1, Annex to Convention (IV) Respecting the Laws and Customs of War on Land: Regulations Concerning the Laws and Customs of War on Land (adopted 18 October 1907, entered into force 26 January 1910) 36 Stat. 2227 TS 539 (hereafter, Hague Regulations); Article 4A(2), Geneva Convention (III) Relative to the Treatment of Prisoners of War (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 135 (hereafter, GC III) (listing subordination, fixed distinctive emblem, carrying arms openly, and conduct in accordance with the laws and customs of war).

insignia,[70] which Parties to armed conflict are obliged to wear to remain "recognizable at a distance".[71] The underlying aim is to make combatants distinguishable from the civilian population, for the latter's protection,[72] but the flipside is to make them amenable to machine perception for lethal targeting.[73] That said, reliance on uniform and insignia alone may lead to distinction failure as combatants may become *hors de combat*;[74] civilians may take the clothing of deceased soldiers and unwittingly put themselves in the crosshairs of a LAWS;[75] or, the enemy may utilise adversarial examples to direct attacking forces towards civilians,[76] in a propaganda war. All three risks counsel in favour of broader criteria and/or multisensory phenomenologies, for a more robust verification of combatant status.[77] One solution may be to combine uniform and insignia with the recognition of specific arms and equipment used exclusively by the enemy.[78] The fact that combatants are legally required to carry their arms openly[79] as a second condition of distinction[80] supports this. A more effective approach may be to detect the 'metallic footprint' and the distinctive 'behaviour and movements', which are a product of military training.[81] Together with uniform and insignia detection this provides a robust three-part criteria,[82] which may also be combined with 'specific arms recognition' when it is desirable to attain a higher

---

[70] See 'Camopedia: The Camouflage Encyclopedia' <http://camopedia.org/index.php?title=Main_Page> accessed 11 August 2018.

[71] Article 1(2), Hague Regulations; Article 4A(2)(b), GC III.

[72] Article 44(3), AP I; CIHL, Rule 106; *AP I Commentary*, ¶ 1578.

[73] Lieutenant Colonel Christopher M. Ford, 'Autonomous Weapons and International Law' (2017) 69 South Carolina Law Review 413, 436.

[74] See 6.5.2.4.

[75] Despite the broader risks associated with wearing military clothing, this is still common in warzones afflicted with harsh winters and poverty amongst the civilian population.

[76] See 2.5.6 and 2.5.7. In this instance, an adversarial example may be subtle patterns embedded in shirts, which are distributed to an unwitting civilian population, and which fool an ATR system into perceiving enemy uniform/insignia.

[77] See also 6.5.2.1.2 on the use of adversarial examples in enemy uniform.

[78] Rao Komar, 'How to Digitally Verify Combatant Affiliation in Middle East Conflicts', *Bellingcat* (9 July 2018) <https://www.bellingcat.com/resources/how-tos/2018/07/09/digitally-verify-middle-east-conflicts/> accessed 11 August 2018.

[79] Article 1(3), Hague Regulations; Article 44(3), AP I; CIHL, Rule 106.

[80] Dinstein (n 6), 52-56 (discussing the seven cumulative Hague and Geneva conditions of lawful combatancy, of which fixed distinctive emblem and carrying arms openly are intended to promote distinction, and are both amenable to ATR. The remaining five are: subordination, conduct in accordance with LOAC, organisation, belonging to a Belligerent Party, and non-allegiance to the detaining power).

[81] William H. Boothby, 'Autonomous Attack – Opportunity or Spectre?' in Terry D. Gill (ed.), *Yearbook of International Humanitarian Law 2013*, Vol. 16 (TMC Asser Press, 2015), 79.

[82] Namely, the *combination* of a) uniform and insignia, b) metallic footprint, and c) the distinctive behaviour and movements, which are a product of military training.

confidence threshold.[83] Along with the legal status of privileged (enemy) combatants, we may expect to see a relatively firm basis for status-based targeting in international armed conflict. Namely, once combatant status is established there are no legal grey areas: active combatants may be attacked based solely on their status, irrespective of the extent of their involvement in hostilities.[84] To be sure, while a small subset of LOAC scholars argues for a so-called 'duty to capture',[85] the overwhelming academic opinion is for status-based targeting;[86] as is the evident legal authority.[87] This clearly supports administrability by precluding the need for a LAWS to undertake any metacognitive conduct-based evaluation, or individualised threat assessment.

However, AP I blurs the 'combatant' category by including other persons who are less amenable to ATR; namely, paramilitary troops and armed police officers who become integrated into the armed forces.[88] That said, the Party integrating such personnel must notify the other Parties to the conflict, to avoid confusion.[89] This will enable the latter to update the ATR of their LAWS, to recognise the relevant uniform and insignia,[90] and the specific arms being used by the paramilitary/police agency.

Even more challenging is the inclusion of guerrilla fighters wearing no uniform or distinguishing sign, and with relaxed rules on the open carriage of their weapons.[91]

---

[83] Or if advances in nanotechnology make metallic footprint redundant. Note that all these approaches would have to pass testing, certification and legal review hurdles before acquisition and deployment.

[84] *Prosecutor* v. *Kordić & Čerkez* (ICTY Appeals Judgment) IT-95-14/2-A (17 December 2004), ¶ 51.

[85] For example, Ryan Goodman, 'The Power to Kill or Capture Enemy Combatants' (2013) 24 The European Journal of International Law 819 (arguing that Articles 35 and 41, AP I, and the general structure, rules and practices of modern warfare all impose restraints on the use of lethal force, and an obligation to utilise the 'least restrictive means', where capture is equally effective and does not endanger attacking forces).

[86] For example, Laurie R. Blank et al., 'Belligerent Targeting and the Invalidity of a Least Harmful Means Rule' (2013) 89 International Law Studies 536 (comprehensively rebutting the so-called 'least restrictive/harmful means' rule, on the basis that a) it ignores the reality of the corporate identity of enemy belligerents, and b) it prevents combatants benefitting from the clarity of presumptions. Namely, armed conflict is a contest between organised belligerent groups, where the objective is to bring the enemy, in a *collective* sense, into complete submission).

[87] Ibid. (providing an extensive examination of positive treaty law and its origins, going back to the Lieber Code through to the Hague Regulations, the GCs and the APs); *Prosecutor* v. *Kordić & Čerkez* (Appeals), ¶ 51.

[88] Article 43(3), AP I; *AMW Manual Commentary*, Rule 10(b)(i), ¶ 3; CIHL, Rule 4.

[89] Article 43(3), AP I; *AMW Manual Commentary*, Rule 10(b)(i), ¶ 3; CIHL, Rule 4.

[90] *AP I Commentary*, ¶ 1683 (affirming that only *uniformed* units of police agencies can be integrated into the armed forces; in line with existing law and State practice, and to avoid confusion).

[91] Article 44(3), AP I (restricting the duty to carry arms openly to the duration of the attack; and to the preliminary phase of deployment in preparation for launching the attack, while being "visible to the adversary").

This is a significant drawback for a LAWS-deploying Party as it undermines the objective dimension of status-based targeting, which is where machines are likely to excel.[92] However, not all States that are expected to field LAWS are Party to AP I, or bound by any (debatable) equivalent rule in customary law.[93] Persistent objectors include the US[94] and Israel,[95] who specifically admonish Article 44(3) and do not recognise it as having customary status,[96] but instead use the traditional categories of combatant found in Article 4A, GC III. Other States that are bound by Article 44(3) have generally expressed three limiting factors, which further narrow the exception and minimise its negative impact on the utilisation of LAWS.[97] In any event, they can always restrict their LAWS deployments to traditional battlefields, where uniformed combatants are the norm, and which will be relatively more amenable to ATR.[98]

---

[92] Dinstein (n 6), 56-57, 64, suggests that the seven conditions of lawful combatancy are onerous for irregular forces, and that the two focused on distinction – distinctive emblem and open carriage of arms – could become alternative rather than cumulative, because "when one is fulfilled the other may be deemed redundant". While this may be true for regular combatants exercising human discretion, a LAWS will need as many objective cues as possible to reliably assess combatant status; thus, it is desirable to consider both as cumulative conditions.

[93] See CIHL, Rule 106 (discussing Article 44(3) and the conduct of negotiations that led to it).

[94] Beginning with Ronald Reagan's Letter of Transmittal to the United States Senate (29 January 1987), reprinted in (1987) 81 American Journal of International Law 910 (specifically objecting to the relaxed rules on irregular forces distinguishing themselves from the civilian population, and the consequent endangering of civilians). See also the authoritative position of the US on AP I in 'Memorandum for Assistant General Counsel (International), Office of the Secretary of Defense, 1977 Protocols Additional to the Geneva Conventions: Customary International Law Implications' (8 May 1986) (identifying the AP I provisions which the US agrees are customary, with a conspicuous absence of Article 44(3)); US Department of Defense, *Law of War Manual* (DoD, 2015; December 2016 Update) (hereafter, *US DoD Manual*), § 4.6.1.2 ("The United States has objected to the way [Article 44] relaxed the requirements for obtaining the privileges of combatant status, and did not ratify AP I, in large part, because of them").

[95] See 'Israel, Statement at the Diplomatic Conference Leading to the Adoption of the Additional Protocols', *Official Records of the Diplomatic Conference on the Reaffirmation and Development of International Humanitarian Law Applicable in Armed Conflicts, Vol. VI* (Federal Political Department, 1978), CDDH/SR 40, ¶ 17.

[96] *US DoD Manual*, § 4.6.1.2 (expressing the US view that Article 44 is not customary law). Even if we do assume customary status, persistent objectors are not bound. See *Fisheries Case (United Kingdom* v. *Norway)* (Judgment) [1951] ICJ Rep 116, 131; Ted L. Stein, 'The Approach of the Different Drummer: The Principle of the Persistent Objector in International Law' (1985) 26 Harvard International Law Journal 457.

[97] (1) The exception is limited to organised resistance movements in occupied territories or in wars of national liberation, (2) 'deployment' refers to any movement towards a place from which an attack is to be launched, and (3) 'visible' includes being detectable via technical means, which is particularly important in relation to LAWS deployments. See the various statements in the *Official Records of the Diplomatic Conference* (n 95), CDDH/SR 40-41.

[98] Following this, if LAWS deployments are found to have positive impacts on civilian risk compared with manned targeting, this may give fresh impetus to calls for extending combatant immunity to non-State actors. See Geoffrey S. Corn, 'Thinking the Unthinkable: Has the Time Come to Offer Combatant Immunity to Non-State Actors?' (2011) 22 Stanford Law & Policy Review 253. The aim would be to incentivise currently 'unlawful combatants' to distinguish themselves as a *quid pro quo*, much to the benefit of civilians and the LAWS-deploying side.

6.5.2.1.2 The Legal Position on Uniforms and Adversarial Examples[99]

Yet, even traditional combatants will not always be a guarantee of adequate distinction. For example, adversarial static may be embedded in military uniform to spoof a LAWS into perceiving civilian clothing; this may amount to no more than a lawful ruse of war if the 'spoofing' Party merely diverts the LAWS away from itself to avoid coming under attack.[100] Indeed, such a move may even be comparable (albeit inversely) to the use of chaff and flares, which has long been practiced by military pilots to divert radar-guided and infrared-guided missiles.[101] On the other hand, if the adversarial static imitates *enemy* uniform and insignia (or flags/emblems)[102] to shield, favour, protect or impede military operations, this would very likely be prohibited under Article 39(2), AP I.[103] Furthermore, if such misuse extends to static-generated uniforms, signs or emblems of the UN or neutral/non-Party States (or civilian clothing patterns) in order to feign protected status; and if this leads to the killing, injuring or capture of LAWS-deploying personnel, it will be deemed to be a perfidy under Article 37(1).[104]

In this regard, Sassóli poses the question as to whether a machine can be 'led to believe' that the person or object before it has protected status, or whether it is possible to 'invite the confidence' of a LAWS – two vital elements of perfidy.[105] Arguably, such anthropomorphic terms cannot directly apply to a LAWS. On the other hand, any manipulation of visual data by an adverse Party that causes a LAWS to hold fire may, by extension, invite the confidence of human combatants who rely on ATR assessments and/or are led by the actions of their machine 'partners'. This may possibly lead those human combatants to believe that the person or object before them enjoys protected status. If such a scenario played out and adverse forces killed, injured or captured attacking forces as a result, perfidy would be very likely to be established.

---

[99] See 2.5.6 on adversarial examples.

[100] Article 37(2), AP I; *AMW Manual*, Rule 113; CIHL, Rule 57.

[101] See the examples of (especially US) State practice under CIHL, Rule 57 <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule57> accessed 11 August 2018.

[102] 'Enemy' in this case being attacking forces.

[103] Article 39(2), AP I; *AMW Manual*, Rule 112(c); CIHL, Rule 62.

[104] Article 37(1), AP I; *AMW Manual*, Rule 111(a)-(b); CIHL, Rule 65.

[105] Sassóli (n 29), 328.

There are limitations to this prohibition: property damage (including damage to the LAWS unit) is not covered by perfidy,[106] even if this does degrade combat capability. However, this may not matter in the case of Article 39(2), which is drafted broadly enough ("impede military operations") to catch property damage or any other degradation to combat capability.[107] The legal difference therefore hinges on whether the adversarial examples centre around enemy uniform and insignia (or flags/emblems); or whether they imitate civilian clothing, or the uniforms, signs or emblems of neutral/non-Party States.

It should be noted that resolving the problem of adversarial examples in uniform is not solely dependent on a legal solution. A more pragmatic approach may be for intelligence analysts to determine – most likely during Phase 2 of the deliberate targeting cycle[108] – how adversarial examples are being utilised by the enemy. Thereon, systems may be trained to recognise specific examples and even specific *kinds* of adversarial patterns in enemy uniform,[109] before being deployed at Phase 5.[110]

6.5.2.2 Civilians and Other Protected Persons

In contradistinction to the above, Article 51(2) prohibits making civilians the *object* of attack,[111] as well as acts or threats of violence for the *primary purpose* of terrorising the civilian population.[112] Paragraph (6) prohibits attacks against civilians by way of *reprisal*,[113] and this is of peremptory importance,[114] given how easily reprisals have in the past been invoked as a pretext for indiscriminate warfare.[115] Arguably, the nature and wording of these prohibitions renders compliance relatively simple: a matter of *ex*

[106] *AMW Manual Commentary*, Rule 111(a), ¶ 7 (noting the inherent limitation in "killing, injuring or capturing").

[107] See 5.3.2.2 on preserving combat capability as a tactical and operational imperative.

[108] See 5.3.1.2.

[109] See 2.5.6.3 on inoculating systems against spoofing.

[110] See 5.3.1.5.

[111] Article 51(2), AP I (emphasis added); *AMW Manual*, Rule 11; CIHL, Rule 1 ("Attacks must not be directed against civilians"); *Nuclear Weapons* AO, ¶ 78 ("States must never make civilians the object of attack"). This general rule is subject to an exception under Article 51(3); on which, see 6.5.2.3.

[112] Article 51(2), AP I (emphasis added); *AMW Manual*, Rule 18; CIHL, Rule 2.

[113] Article 51(6), AP I (emphasis added). See also CIHL, Rule 145.

[114] *AP I Commentary*, ¶ 1984.

[115] Ibid., ¶ 1982 (referring to the "countless civilian victims" during World War II). But see CIHL, Rule 145 (restating the customary norm and conditions under which reprisals may be permissible).

ante programming and appropriate deployment,[116] which should pose no difficulty for commanders acting in good faith.[117]

The issue differs when considering the general civilian protection afforded in Paragraph (1),[118] Article 48[119] and in customary law.[120] None of these presume deliberate targeting on the part of human participants, and all of them may be violated when there is distinction failure on the part of the machine, *if this would not occur in a counterfactual manned targeting scenario*.[121] Yet, even here a LAWS is – at least in more traditional battlefield contexts – arguably capable of respecting civilian status by recognising "any non-positively identified person" as a civilian.[122] This kind of programming is consistent with the negative definition of 'civilian' in Article 50(1), AP I,[123] and it would entail a technical prohibition on targeting any person *not* satisfying the three-/four-part criteria discussed above. Thus, far from being vague and non-executable in machine code, it simply requires programming the inverse of the status-based criteria.[124]

[116] Schmitt (n 63).

[117] *AP I Commentary*, ¶ 2198.

[118] Article 51(1), AP I (noting subsequent paragraphs are in addition to the general protection enjoyed by civilians).

[119] Article 48, AP I (noting simply that Parties to a conflict "must at all time distinguish between civilians and combatants").

[120] See restatements in CIHL, Rules 1 and 6 (the former replicating the wording of Article 48, AP I; the latter restating the rule that civilians are generally "protected against attack").

[121] Schmitt (n 19), 12; Sassóli (n 29), 320; *AMW Manual Commentary*, Rule 39, ¶ 4.

[122] Schmitt (n 63). 'Respect' in this context simply means to refrain from attacking civilians, and is not meant to be an anthropomorphism.

[123] Article 50(1), AP I; *AMW Manual Commentary*, Rule 11, ¶ 6; CIHL, Rule 5; *Prosecutor* v. *Blaškić* (ICTY Trial Judgment) IT-95-14-T (3 March 2000), ¶ 180.

[124] Cf. Noel E. Sharkey, 'The Evitability of Autonomous Robot Warfare' (2012) 94 International Review of the Red Cross 787, 789 (arguing that such a negative formulation "does not provide a definition that could give a machine with the necessary information").

Other persons who must be both respected and protected include medical,[125] religious[126] and humanitarian relief personnel,[127] amongst others.[128] On the one hand, these specific categories of persons may require no further programming efforts than those that will be afforded to civilians, as these persons will also satisfy the inverse of the status-based criteria for combatants. This is helpfully reinforced by the restriction of medical personnel to "light individual weapons"[129] in AP I[130] and in the *2016 GC I Commentary*,[131] to avoid the perception that they are equipped to commit (outside their humanitarian duties) acts harmful to the enemy.[132] Namely, as permissible and 'restricted' arms are all amenable to object recognition,[133] these rules will potentially support respect for medical personnel by LAWS-deploying forces.

On the other hand, respect may be bolstered by programming additional (positive) forbidding criteria when these persons bear machine-perceptible signs. Specifically, medical and religious personnel are required to wear a water-resistant armlet bearing the emblem of the Red Cross or Red Crescent, to denote protected status to attacking forces.[134] Moreover, when carrying out their duties in a battle area, these persons shall

---

[125] Articles 24 and 25, Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 31 (hereafter, GC I); Article 36, Geneva Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 85 (hereafter, GC II); Article 20, Geneva Convention (IV) Relative to the Protection of Civilian Persons in Time of War (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 287 (hereafter, GC IV); Article 15(1), AP I; *AMW Manual*, Rule 71; CIHL Rule 25.

[126] Article 24, GC I; Article 36, GC II; Article 15(5), AP I; *AMW Manual*, Rule 71; CIHL Rule 27.

[127] Article 71(2), AP I; Article 7(2), Convention on the Safety of United Nations and Associated Personnel (adopted 9 December 1994, entered into force 15 January 1999) 2051 UNTS 363; *AMW Manual*, Rule 102(a); CIHL, Rule 31.

[128] See, for example, Articles 26 and 27, GC I, on personnel of National Red Cross and other Voluntary Aid Societies, and Societies of neutral countries, respectively.

[129] See Heather Brandon, 'Joint Series: Restricting Medical Personnel, Units, and Transports to 'Light Individual Weapons'', *Intercross Blog* (16 February 2017) <http://intercrossblog.icrc.org/blog/joint-series-restricting-medical-personnel-units-and-transports-to-light-individual-weapons> accessed 17 August 2018.

[130] Article 13(2)(a), AP I; *AMW Manual*, Rule 74(c)(i).

[131] Knut Dörmann et al. (eds.), *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field* (CUP, 2016), ¶¶ 1864 and 1874 (in relation to military medical personnel).

[132] Ibid., ¶ 1868.

[133] Ibid., ¶¶ 1864 and 1868 (noting that pistols and standard-issue rifles are acceptable, but man-portable and anti-tank missiles and crew-served machine guns are not). See also *AMW Manual Commentary*, Rule 74(c)(i), ¶ 3.

[134] Articles 38-41, GC I; Articles 41-42, GC II; Article 4 and 5, Annex I to Protocol I Additional to the Geneva Conventions of 1949: Regulations Concerning Identification (as amended on 30 November 1993, entered into force 1 March 1994) (hereafter, Amended Annex I); *AMW Manual*, Rule 72(a); CIHL, Rule 30.

– as far as possible – wear headgear and clothing that also bears the distinctive emblem,[135] and they may (should) use materials that make the emblem recognisable by technical means of detection.[136] These should further increase the likelihood of reliable machine perception and the application of forbidding criteria by a LAWS.

So far, there appears to be a relatively clear textual basis for distinction between categories of persons, which LAWS may be expected to satisfy in at least some circumstances. However, in contemporary conflicts sub-categories often appear, which complicate the distinction task.

6.5.2.3 Civilians Not Protected from Direct Attack

In particular, civilians may take a direct part in hostilities (DPH) and, *for such time* that they do, they become targetable.[137] This temporal element complicates matters because to be liable to attack, a civilian must act on a 'spontaneous, sporadic or unorganised basis'[138] and must, in the view of the ICRC, cumulatively meet its *threshold of harm* with *direct causation* and a *belligerent nexus*.[139] Moreover, measures taken to prepare for a specific act of DPH, as well as deployment to and from the location of that act, also qualify as DPH.[140] On the other hand, the ICRC considers there to be a 'revolving door', whereby suspension of civilian protection lasts only as long as the person engages in DPH,[141] even if there are persistently recurrent cycles.[142] While this legal view is often disputed,[143] it is generally acknowledged that the factual

---

[135] Article 5(4), Amended Annex I.

[136] Article 5(3), Amended Annex I; *AMW Manual Commentary*, Rule 72(b), ¶ 2 (suggesting thermal tapes).

[137] Article 51(3), AP I; *AMW Manual*, Rule 28; CIHL Rule 6.

[138] International Committee of the Red Cross (ICRC), *Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law* (ICRC, 2009), 34.

[139] Ibid., 46-64 (discussing the 'Constitutive Elements' of DPH).

[140] Ibid., 65-68.

[141] Ibid., 70-73.

[142] Ibid., 44.

[143] See, for example, William H. Boothby, 'Direct Participation in Hostilities – A Discussion of the ICRC Interpretive Guidance' (2010) 1 International Humanitarian Legal Studies 143, 162 (arguing that repeated and persistent DPH is a reliable indicator as to future conduct, and that a persistent participator should remain targetable until he renders an "overt and unambiguous act of renunciation"); Michael N. Schmitt, 'The Interpretive Guidance on the Notion of Direct Participation in Hostilities' (2010) 1 Harvard National Security Journal 5, 36 (arguing that 'for such time' should extend "as far before and after a hostile action as a causal connection existed"); *US DoD Manual*, §5.8.4.1 (laying down the official US position, that persons DPH remain targetable until they have "permanently ceased participation in hostilities").

circumstances giving rise to DPH in the first instance are not always objectively discernible.[144]

This creates a conduct-based targeting challenge that will be very difficult for near-term LAWS to meet. Specifically, ATR systems will find it very difficult to recognise offensive behaviour from a civilian, with no other tangible cues.[145] On the other hand, three potential solutions have been suggested. First, Henderson, Keane and Liddy argue that *some* DPH indicators currently applied by human decision-makers are relatively tangible, and may be programmed into a LAWS.[146] These include whether an individual is openly armed; his proximity to the fighting and/or other military equipment; and the direction and manner of his movement.[147] So long as each characteristic is appropriately weighted, it is conceivable that a *combination* of such criteria pointing in the same direction might be a strong indicator of a civilian undertaking DPH.[148] However, this approach seems plausible only in a limited range of circumstances, with much scope for erroneous targeting. Namely, it ignores the near-infinite combinations of relevant cues, the metacognitive approach of human soldiers in situations of uncertainty[149] and their reliance on "gut feeling"[150] *versus* the deterministic response of a robot.

---

[144] Michael N. Schmitt and Eric W. Widmar, 'On Target: Precision and Balance in the Contemporary Law of Targeting', (2014) 7 Journal of National Security Law & Policy 379, 390 ("The status of an individual can sometimes be unclear…consider…civilians sitting on a hillside overlooking a commonly used helicopter landing zone. Without additional intelligence indicating they are being used as an early warning system…IHL requires them to be treated as civilians and protected from attack"). See also the examples of 'hostile intent' and 'hostile act' illustrated by Gaston in Chapter 5, n 146 therein.

[145] Markus Wagner, 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems' (2014) 47 Vanderbilt Journal of Transnational Law 1371, 1392-93 (commenting on the contextual reasoning required to accurately assess DPH, and to distinguish it from innocuous activity or lawful self-defence). See also 2.5.5.2, especially the *kirpan* example at n 296.

[146] Ian S. Henderson, Patrick Keane and Josh Liddy, 'Remote and Autonomous Warfare Systems: Precautions in Attack and Individual Accountability' in Jens David Ohlin (ed.), *Research Handbook on Remote Warfare* (Edward Elgar, 2017), 346-47.

[147] Ibid.

[148] A remaining legal difficulty will be to set the required degree of confidence and an acceptable failure rate for a LAWS to determine this.

[149] Boothby (n 31), 149 (noting that in ambiguous situations of DPH, the soldier must "take into account all information that is reasonably available to him…including any indication that there may be as to the *reliability* of information from that source") (emphasis added).

[150] Robin Geiss, *The International Law Dimension of Autonomous Weapons Systems* (Friedrich Ebert Stiftung Study, October 2015), 14 (also noting the difficulty of configuring machine perception for equivocal combat situations).

Second, Ford considers a 'narrow deployment' approach, focusing on the common insurgency practice of emplacing an improvised explosive device along a road. The author argues that this is both amenable to machine perception[151] and it potentially justifies lethal attack by a LAWS, in a way compatible with the ICRC's *Interpretive Guidance*.[152] However, this kind of deployment may also be subject to targeting error. For example, a LAWS may detect a person with explosive chemical signatures on a construction site, and open fire on civilians undertaking innocuous commercial activity; though there are also control mechanisms to mitigate this risk.[153]

Finally, perhaps the most commonly-cited approach is Arkin's 'conservative use of lethal force' concept.[154] This argues that robots do not necessarily have a self-preservation instinct, thus can be programmed to hold fire on all civilians *until fired upon*.[155] Arguably, opening fire on a LAWS goes beyond mere 'hostile intent'[156] and may be regarded as *prima facie* evidence of a 'specific hostile act', which is the very essence of DPH.[157] Moreover, with the use of optical/acoustic detection systems, gunfire is easily recognisable by a robot[158] and should *legally* permit a defensive lethal response. However, even this approach is problematic. As mentioned in 2.5.4.3, gunfire is merely 'target indication', which in most cases will require a 'tipping and cueing' of sensors and further (automatic or controlled) processing before it can progress to full 'target identification'. Thus, relying on gunfire alone may lead to numerous other problems, including:

---

[151] Ford (n 73), 438 (referring to radar-based human detection and spectrographs that can detect the chemical signature of explosives).

[152] Ibid. (noting that a LAWS would be able to track the insurgent while deploying to and/or from the location of emplacement, and wait until no other civilians are present before lethally engaging him).

[153] Ibid., 438-39 (suggesting: (1) deploying appropriately sophisticated LAWS, (2) spatio-temporal restrictions, (3) system updates on specific persons undertaking DPH, and (4) operator control. Many of these implicate the deliberate targeting process, during which risks can be further mitigated with intelligence-gathering and deployment only in circumstances where lawful activities with overlapping indicators have been ruled out).

[154] Ronald C. Arkin, *Governing Lethal Behaviour in Autonomous Robotics* (Chapman & Hall/CRC, 2009), 46.

[155] Michael N. Schmitt and Jeffrey S. Thurnher, 'Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict' (2013) 4 Harvard National Security Journal 231, 264.

[156] This ROE term was rejected in the *Interpretive Guidance* for being too context-specific, hence "unhelpful, confusing or even dangerous" to apply as general guidance for defining DPH: ICRC (n 138), 52.

[157] Ibid., 43-45. For a definition of both 'hostile intent' and 'hostile act', see NATO Military Committee, *NATO Rules of Engagement*, MC 362/1 (NATO HQ, 30 June 2003), Appendix 1, Annex A, ¶¶ 3-5.

[158] See 2.5.4.3 on target indication.

- The risk of 'shoot and scoot', where insurgents open fire from areas of civilian concentration, before fleeing to confuse the adversary.[159] This may become a common 'baiting tactic' if LAWS are deployed in urban areas or equipped with indirect fires,[160] and it may lead them to return fire into civilian areas, even though a metacognitive human may have had cause to hesitate and reassess.
- The risk that the insurgent is using a human shield, and the likelihood that a LAWS will return fire and kill the latter. While the legal status of human shields is controversial,[161] there is near-consensus that *involuntary* human shields retain their protected status.[162] In which case, precautions in attack must be taken, and any expected harm to them must be fully factored into the proportionality assessment.[163]
- In the most chaotic situations, there is the broader risk of civilians being caught in the cross-fire, as well as fratricide.[164]

Consequently, while narrow conditions may exist where civilians taking DPH are amenable to autonomous attack, the risk of unforeseen circumstances, elusive behaviour and consequent distinction failure is arguably too great to give LAWS target engagement authority in a DPH setting. Thus, near-term LAWS deployments will be better-suited to traditional battlefields, where enemy combatants offer a clearer basis for distinction.

6.5.2.4 Persons Hors de Combat

Yet, even in such battlefields, there may remain the problem of systems not recognising when combatants become persons *hors de combat*,[165] and thus protected

---

[159] Michael N. Schmitt, 'The Principle of Distinction and Weapon Systems on the Contemporary Battlefield' (2008) 7 Connections 46, 54.

[160] Ibid., 55; Colonel Richard Jackson, 'Autonomous Weaponry and Armed Conflict', *ASIL Panel Discussion* (10 April 2014) <https://www.youtube.com/watch?v=duq3DtFJtWg> accessed 10 May 2018.

[161] Michael N. Schmitt, 'Human Shields in International Humanitarian Law' (2009) 47 Columbia Journal of Transnational Law 292. For a brief survey of the various views, see Boothby (n 31), 137-39; Schmitt and Widmar (n 144), 388-89.

[162] Boothby, ibid., 136-37; Ian Henderson, *The Contemporary Law of Targeting: Military Objectives, Proportionality and Precautions in Attack under Additional Protocol I* (Martinus Nijhoff, 2009), 215.

[163] Ibid. Again, this implicates the need for controlled processing and metacognitive thinking at the point of trigger-pull/weapons release.

[164] For a dramatic account of the "messy chaos of war" where human soldiers over-reacted and inadvertently shot each other in response to gunshots, see Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (Norton, 2018), 253-55 (specifically linking this to the expected risk of poor performance by a LAWS deployed in similar circumstances).

[165] Sassóli, (n 29), 327.

from direct attack.[166] This may occur in one of three ways: a) capture by friendly forces, b) clearly expressing an intention to surrender, or c) incapacitation, hence an inability to defend oneself.[167] The first of these is not relevant to LAWS, as captured personnel are under the control of the LAWS-deploying side.[168] The second may be simple or difficult, depending on context and circumstances. For example, in demilitarised zones some basic surrender recognition capabilities currently exist,[169] which may be complemented by more recent developments in deep learning for emotion-reading.[170] In active combat situations, there are a number of other potential (albeit imperfect) solutions,[171] the most robust being a restriction of LAWS deployments to particular operational environments;[172] for example, combat between armoured vehicles or submarines, where established conventions on surrender are amenable to machine perception.[173] In the case of anti-personnel targeting, there is also the fact that persons *hors de combat* – be that via surrender or incapacitation – will clearly cease any military-style 'behaviour and movements', and this *may* negate the three-/four-part criteria for detecting active combatant status.[174]

However, in more difficult surrender contexts, as well as incapacitation that is not amenable to machine perception,[175] the technical challenges will bring into play some important legal questions. The wording of Article 41(1), AP I, is particularly important here, where it prohibits the targeting of anyone "who is recognized or who, in the

[166] Article 41(1), AP I; *AMW Manual*, Rule 15(b); CIHL, Rule 47.

[167] Article 41(2), AP I; *AMW Manual*, Rule 15(b); CIHL, Rule 47.

[168] See also *AMW Manual Commentary*, Rule 15(b), ¶ 3 (excluding capture for being irrelevant to aerial warfare).

[169] See 'Samsung Techwin SGR-A1 Sentry Guard Robot', *GlobalSecurity.org* (7 November 2011) <http://www.globalsecurity.org/military/world/rok/sgr-a1.htm> accessed 17 August 2018 (detailing the now-retired *Samsung SGR-A1* sentry robot, which was equipped with gesture recognition technology, to recognise 'arms held high' as a sign of surrender).

[170] See 2.5.2, and especially (notes and text accompanying) nn 204-205 within that.

[171] Robert Sparrow, 'Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender' (2015) 91 International Law Studies 699, 712-18 (discussing four control mechanisms: humans-in-the-loop; restriction to anti-material targeting and/or non-lethal munitions; surrender beacons; and restriction of LAWS to particular operational environments).

[172] Ibid., 717-18.

[173] Ibid., 718 (noting the reversing of turrets and opening the hatch on a tank; and submarines using an 'underwater telephone', or rising to the surface and waving a white flag).

[174] Or, they can throw aside all weapons and raise their arms, as suggested in *AMW Manual Commentary*, Rule 15(b), ¶ 7.

[175] For example, where the combatant initially satisfies the three-/four-part criteria and falls into the crosshairs, but becomes *hors de combat* in a way that may be mistaken for taking cover. See ibid., ¶ 6.

circumstances, *should be recognized* to be *hors de combat*".[176] According to Boothby, this means that if an alternative and reasonably available means or method of attack would permit such recognition, the "should be recognized" criterion is satisfied, and if a LAWS erroneously proceeds with an attack, the rule is violated.[177] This interpretation of Article 41(1) assumes that the requirement of 'feasibility' under Article 57(2)(a)(i)[178] sets the correct standard and, if accepted, would mean that limitations of ATR technologies will not, in themselves, afford an excuse for failing to comply with the principle of distinction. Consequently, commanders will have to consider very carefully their deployment options, even in simple and remote battlefields.

However, there is a compelling counter-argument with Henderson, Keane and Liddy contending that Article 41(1) itself sets the correct standard for determining *hors de combat*.[179] Under their approach, the legal issue is not whether an alternative and reasonably available weapon system would have permitted accurate recognition; but whether, *based on the actual weapon system employed*, a person should have been recognised as being *hors de combat*.[180] This would appear to be consistent with State practice: means and methods of warfare have long involved indirect fires[181] and, since the 1960s, a range of other 'beyond-visual-range' (BVR) engagements, particularly in air combat.[182] None of these assist attackers in determining whether persons to be engaged are *hors de combat*, yet they continue to be routinely deployed with no legal difficulty. Moreover, as the *AMW Manual Commentary* explains, combatants must *effectively* communicate their intention to surrender. If they do not, and if attackers conducting a BVR engagement remain unaware of their intention to surrender, the attack may lawfully proceed; so long as the lack of knowledge on the part of the attackers is reasonable in the circumstances.[183]

---

[176] Article 41(1), AP I (emphasis added).

[177] Boothby (n 64), 109.

[178] See 7.3.2.1.

[179] Henderson, Keane and Liddy (n 146), 348.

[180] Ibid.

[181] NR. Jenzen-Jones (ed.), *Indirect Fire: A Technical Analysis of the Employment, Accuracy, and Effects of Indirect-Fire Artillery Weapons* (ARES, January 2017), 15-58 <https://www.icrc.org/en/document/indirect-fire-technical-analysis-employment-accuracy-and-effects-indirect-fire-artillery> accessed 17 August 2018 (detailing the historical background, purpose and types of indirect-fire systems, and issues relating to their employment).

[182] John Stillion, *Trends in Air-to Air Combat: Implications for Future Air Superiority* (CSBA, 2015) <http://csbaonline.org/uploads/documents/Air-to-Air-Report-.pdf> accessed 17 August 2018. See also *AMW Manual Commentary*, Rules 7(c), ¶ 3 and 13(b), ¶ 8.

[183] *AMW Manual Commentary*, Rule 15(b), ¶ 5.

In view of this, where the sensory limitations of a LAWS cause it to fail to detect surrender or incapacitation, this should not in itself render an attack unlawful. To effectively communicate surrender to attacking forces, the burden is on surrendering forces to communicate with the forces conducting an autonomous attack; even if this requires contacting other forces, which can pass the information to the relevant commander/WO in good time.[184]

The issue is even clearer in the ICRC's restatement of customary law, which simply prohibits attacks on "persons who *are* recognized as *hors de combat*", with no alternative 'should be recognised' criterion.[185] This is likely to be the default legal position for States not Party to AP I, such as the US and Israel.

6.5.3 Objects

As far as objects are concerned, 'military objective' is defined in Article 52(2), AP I, as:

> [T]hose objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage.[186]

One of the most heavily debated provisions in AP I,[187] this is often referred to as a 'two-pronged test' in that it comprises two cumulative criteria, as indicated by the conjunctive "and".[188] First, there is the "effective contribution to [the enemy's] military action" (ECMA) by reference to the "nature, location, purpose or use";[189]

---

[184] Ibid.

[185] CIHL, Rule 47 (emphasis added).

[186] Article 52(2), AP I; *AMW Manual*, Rule 1(y). See also CIHL, Rule 8.

[187] Stefan Oeter, 'Means and Methods of Combat' in Dieter Fleck (ed.), *The Handbook of International Humanitarian Law* (3rd ed., OUP, 2013), 169 (noting the hostility of Western militaries to the apparently restrictive drafting, which limits the categories of objects that can be legitimately attacked). Article 52(2) has also given rise to controversies regarding war-sustaining capabilities *versus* harmful reverberating effects. See W. Hays Parks, 'Air War and the Law of War' (1990) 32 Air Force Law Review 1, 135-45; Cf. Judith Gail Gardam, 'Proportionality and Force in International Law' (1993) 87 American Journal of International Law 391, 404-10.

[188] Henderson (n 162), 51; Horace B. Robertson, Jr, 'The Principle of the Military Objective in the Law of Armed Conflict' (1997) 8 United States Air Force Academy Journal of Legal Studies 35, 48.

[189] Despite the apparent 'closed-list approach' and use of the conjunctive "or", which normally indicates alternatives, these four are merely *key* sub-criteria used to guide targeting decisions. They are neither the sole considerations nor are they mutually exclusive. Indeed, it may be necessary to examine more than one category in determining whether there is a *definite* military advantage in the circumstances

second, the "definite military advantage" (DMA) to the attacker, to be assessed "in the circumstances ruling at the time". Both criteria must be fulfilled in light of their qualifiers,[190] though there seems to be no consensus on timing: some argue that both conditions must be *simultaneously* present;[191] others seeing no temporal aspect;[192] yet, others taking a 'middle-ground' with a condition of reasonableness.[193]

An accurate understanding of the definition of military objective and its application is indispensable, not only for commanders to know which objects they can legitimately attack, but also for ensuring humanitarian protection. This is because 'civilian objects', which are immune from direct attack, are defined in the negative.[194] As a general proposition, it is assumed that the broader and/or more concrete the application of each component of the definition, the more amenable it will be to algorithmic determination in the 'narrow loop'.

6.5.3.1 'Nature' and 'Location'

The *AMW Manual Commentary* explains that an object is a military objective by **nature** when its "*inherent* characteristic or attribute" contributes to military action".[195] Examples include military aircraft (including unmanned); military vehicles (excluding medical transports); missile batteries and other weapons; military equipment, fortifications, facilities and depots; warships; and ministries of defence and armaments

---

ruling *at the time of attack*. See Gary D. Solis, *The Law of Armed Conflict: International Humanitarian Law in War* (2nd ed., CUP, 2016), 510. This chimes with Henderson (n 162), 54, who sees the four sub-criteria not as words of limitation, but as a *test* for determining what is a military objective, instead of having a more restrictive list of specific objects. Accordingly, if a true ECMA arises other than through "nature, location, purpose or use", the author argues that there is unlikely to be any objection to an attack solely on that basis. Contrast the drafting of Article 56(1).

[190] Boothby (n 31), 102 (arguing that listing specific examples of military objectives is liable to cause confusion and arbitrariness, and that the real issue is whether an object actually fulfils the Article 52(2) definition *at the time of the decision to attack*. In this regard, the author is consistent with Henderson, ibid.).

[191] *AP I Commentary*, ¶ 2018.

[192] Marco Sassóli, 'Legitimate Targets of Attack Under International Humanitarian Law', *Harvard Program on Humanitarian Policy and Conflict Research, Background Paper* (2003), 7 (omitting the temporal aspect and merely stating that an object must "cumulatively fulfil [the] two criteria").

[193] Henderson (n 162), 52 (arguing that simultaneous fulfilment is not required (too strict), but mere cumulative fulfilment is not enough (too lax), and illustrating this with the example of a computer system used for long-term planning of military operations).

[194] Article 52(1), AP I, defines civilian objects as "all objects which are not military objectives as defined in paragraph 2". See also *AMW Manual*, Rule 1(j); CIHL Study, Rule 9. The *AP I Commentary*, ¶ 2012, justifies this with the fact that there are more civilian objects than military objectives. Arguably, it is also justifiable as being consistent with the rule of 'doubt' in Article 52(3), and with the approach to defining civilian in Article 50(1).

[195] *AMW Manual Commentary*, Rule 22(a), ¶ 1 (emphasis added)

factories. Crucially for LAWS, the military characteristics of these objects are *non-changeable*, meaning that they "*always* constitute lawful targets during armed conflict…even when not in use".[196] This is further reinforced, both in the *AMW Manual Commentary*[197] and by academic opinion.[198] If accepted, it eliminates the need for context-based evaluation at a given point in time, thus facilitating *ex ante* programming[199] and deployment for both a TS and TLC.[200] The *AP I Commentary* appears to take an even broader view, simply defining military objective by nature as "all objects directly used by the armed forces", before providing a relatively short illustrative list;[201] again, facilitating *ex ante* programming. Moreover, given the technical features of ATR explained in 2.5.4.1, these robust definitions render such objects highly machine-perceptible via a quantitative assessment of inherent, non-changeable and easily-recognisable characteristics, like image, size, shape, sound, heat, velocity and material content.[202] The reliability is further bolstered in the case of cooperative targets, which emit signals that can be easily detected by passive sensors;[203] and by drawing on stationary and/or moving target indication.[204]

Conversely, there is a body of academic opinion, which maintains that a military *object* is not a military *objective* by default, and that the former are only targetable if they independently meet the two-pronged test of the latter.[205] In this sense, such opinion diverges from the approach taken in the *AMW Manual Commentary*; not in relation to ECMA *per se*, but in relation to the DMA qualifying the ECMA. Given the two-pronged, cumulative drafting of Article 52(2), it is submitted that this latter view must be correct, with the overall effect that military objects by 'nature' become relatively

---

[196] Ibid.

[197] Ibid., ¶ 3 (stating, "[t]heir distinctive feature is that they qualify as military objectives by nature *in all circumstances*") (emphasis added).

[198] For example, Robertson (n 188), 49 (arguing "[s]ome objects, 'by their nature', are military objectives and *remain so at all times, regardless of their location or use*") (emphasis added).

[199] Dinstein – one of the key architects of the *AMW Manual* – similarly defines 'nature' by reference to the "intrinsic [military] character" of an objective, before suggesting a more detailed (non-exhaustive) list. See Dinstein (n 6), 110-11; Yoram Dinstein, 'Legitimate Military Objectives under the Current Jus in Bello' (2002) 78 International Law Studies 139, 146-47.

[200] For fixed and moving objects, respectively; hence, subject to relatively more careful programming of target parameters in the case of TLC.

[201] *AP I Commentary*, ¶ 2020 (listing weapons, equipment, transports, fortifications, depots, and buildings used by armed forces, such as staff headquarters and communications centres).

[202] See 2.5.4.1, especially on how multisensory phenomenologies and cross-cueing aids the process.

[203] See 2.5.4.2.

[204] See 2.5.4.3.

[205] For example, Boothby (n 31), 103; Henderson (n 162), 51 and 55.

less amenable at autonomous attack. Conversely, as will be seen below in 6.5.3.3, the DMA alone rarely negates the definition of military objective.

An object is a military objective by **location** when its geographical location makes an ECMA, irrespective of its nature,[206] use, or even purpose.[207] This includes bridges situated in militarily strategic areas, or even a specific area of land *en masse*,[208] where it is important for military operations to seize that location, to deny the enemy from seizing it, or to force the enemy to retreat from it.[209] Accordingly, attacking a location is only lawful under certain circumstances,[210] as every plot of land is unique and may offer a shifting and contextual value for the enemy's military action. This calls for deliberative human input in the determination of which *specific* locations to target.

Concretely, this means that – unlike military objects by 'nature' that can be engaged in TLC – attacking a 'location' will almost certainly have to be a TS. It will need to benefit from the controlled processing and metacognitive thinking that goes into the deliberate or dynamic targeting cycle, leaving the weapon system to act autonomously only in relation to the *timing of the attack* and (potentially) the *munition selected*. It is also likely that, in line with precautionary measures,[211] commanders will have to demarcate the *smallest area of land* consistent with the requirements of military necessity,[212] and the one whose military utility (for the enemy) is *least likely to erode* during the time of deployment. However, contrary to the unsupported assertion of the

---

[206] *AP I Commentary*, ¶ 2021 (giving the canonical example of a bridge that is/may become part of a militarily important route).

[207] *AMW Manual Commentary*, Rule 22(b) (referring to mountain passes that may be blocked in case the enemy needs to retreat; attacking high ground to blind the enemy; or destroying natural cover to deprive them of an observation point. In all cases, the attack may be to safeguard the attacker's operations and diminish the enemy's options, irrespective of actual current or intended future use by the enemy).

[208] Marco Sassóli, Antoine A. Bouvier and Anne Quintin, *How Does the Law Protect in War?: Cases, Documents and Teaching Materials on Contemporary Practice in International Humanitarian Law, Vol. I* (3rd ed., ICRC, 2011), Chapter 9, 5.

[209] *AP I Commentary*, ¶ 2021.

[210] Ibid., ¶ 2025(a) (noting country statements and declarations, which mentioned "circumstances" as well as "location" as relevant factors that can turn a specific area of land into a military objective); Dinstein (n 6), 115; (n 199), 150 ("There must be a distinctive feature turning a piece of land into a military objective, e.g., an important mountain pass or defile, a specific hill of strategic value, a bridgehead or a spit of land controlling the entrance of a harbour"). See also *AMW Manual Commentary*, Rule 22(b) and its accompanying commentary (summarised at n 207, above).

[211] See 7.3 and especially 7.3.2.2 on the choice of target that minimises civilian risk.

[212] *AP I Commentary*, ¶ 2026 (stating that an area of land being targeted "can only be of a limited size" and this is cited with approval by Henderson (n 162), 56; presumably to avoid speculative attacks and to retain a *definite* military advantage).

AP I Commentary,[213] a location-based military objective is unlikely to be restricted to the immediate combat area,[214] especially when areas outside the contact zone are typically used as logistical routes.[215] There is no obvious reason why this would be any different in the case of an autonomous attack.

Once these legal boundaries are applied and integrated into the targeting process, location becomes the most amenable to machine perception of the four ECMA sub-criteria. Unlike the determination of an object's 'nature', which calls for stochastic reasoning, a location is objectively ascertainable via the Global Positioning System (GPS).[216] Even in denied areas, where GPS guidance systems may be ineffective or vulnerable to hacking, a LAWS will still be able to operate reliably via electro-optical/infrared scene-matching.[217]

To summarise, LAWS fitted with appropriate ATR and guidance systems are indeed capable of being deployed in compliance with the nature and location sub-criteria. This is helped by the fact that the *effective* contribution to military action need not be critical or even significant for an object to qualify as a targetable military objective;[218] so long as it does *in fact* contribute to the enemy's military action.[219] Arguably, the binary nature of this condition supports the application of presumptions over context-based evaluation, thus making it more likely that *ex ante* programming and machine perception will operate in line with legal requirements.

6.5.3.2 'Purpose' and 'Use': The Problem of 'Dual-Use' Objects

In contrast, the last two ECMA sub-criteria – 'use' and 'purpose' – are more difficult to assess, both for human soldiers and, even more so, for LAWS. At their core is the

---

[213] *AP I Commentary*, ¶ 2026 (asserting that the 'location' concept "is only valid in the combat area", though with no further discussion and citing no authority for this).

[214] Henderson (n 162), 56-57 (pointing out the lack of support for this limitation in the *AP I Commentary*, and that it is conspicuously absent in both the negotiating history of AP I and in State practice. Furthermore, there is no discernible need to read such a limitation into 'location' when the same is not applied to 'nature', 'use' or 'purpose').

[215] This point is particularly important as it accords with the idea of offering an ECMA to the enemy (if preserved) and a DMA to the attacking party (if destroyed, captured or neutralised).

[216] See 2.5.4.4. Crucially, GPS guidance systems can also *prevent* attacks on locations placed on a 'no-strike' list, such as fixed medical units, non-defended localities and demilitarised zones under Articles 12, 59 and 60, AP I, respectively. See also 6.5.3.4.

[217] See 2.5.4.5.

[218] Schmitt and Widmar, (n 144), 392.

[219] *AMW Manual Commentary*, Rule 1(y), ¶ 4.

fact that both concepts involve *dual-use objects*; namely, those that "simultaneously serve both the military and the civilian population of the enemy".[220] Such an object, "on the face of it, is civilian in nature…but subsequently becomes a lawful target as a result of conversion to military use".[221] Crucially, with simultaneous military and civilian use, an attack can only proceed subject to the principle of proportionality,[222] which presents an extra layer of complexity for a LAWS, with perhaps a greater need for controlled processing during the targeting cycle.[223]

While not strictly a legal term, dual-use objects give rise to two distinct factual problems. First, attacks on these objects often have a more perilous effect on civilians, either because the latter are more likely to be present and/or because these attacks tend to inflict damage or pose dangers to them that continue for long periods;[224] though, this is more a *policy* and legal *proportionality* concern. Second, and more pertinent to the principle of distinction, both sub-criteria are highly malleable, especially during hostilities; hence, they demand that greater attention be paid to the adjectives "effective" and "definite", to ensure the object being targeted really has acquired the legal status of 'military objective'.[225]

That said, it was noted above that the *effective* contribution need not be critical or even significant; so long as an ECMA does in fact exist. Furthermore, ECMA does not presuppose a *direct* connection with combat operations, as is implied in Article 51(3) regarding persons.[226] Thus, Article 52(2) can make a civilian object targetable through 'use' or 'purpose' that is only *indirectly* related to military action; again, so long as it makes an *effective* contribution.[227] This relative breadth of the ECMA concept can, in

---

[220] Dinstein (n 6), 120. See also *AP I Commentary*, ¶ 2023.

[221] *AMW Manual Commentary*, Rule 22(d), ¶ 1.

[222] Ibid., Rule 22(d), ¶ 7; *AP I Commentary*, ¶ 2023.

[223] See 7.2.

[224] Boothby (n 31), 104-05 (arguing that such 'reverberating' effects must also be considered, both in the proportionality analysis and in applying precautionary measures. This is to get a more accurate measure of the expected collateral damage, for comparison with the concrete and direct military advantage anticipated. See also Gardam (n 187), 404-10, on the attacks on civilian infrastructure during the Gulf War).

[225] Boothby, ibid., 105.

[226] Michael Bothe, Karl Josef Partsche and Waldemar A. Solf, *New Rules for Victims of Armed Conflict: Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949* (Martinus Nijhoff, 1982), 324. Namely, Article 51(3) makes civilians targetable only when they take a *direct* part in hostilities.

[227] Ibid. Hence, the possible inclusion of factories that 'merely' contribute to the war effort, for example, by producing soldier's uniform or parts for military vehicles.

some circumstances, make it more amenable to autonomous application, although the bigger picture is that the determination of use and purpose will generally need a greater input of deliberative human reasoning.

'Use' means the enemy is *presently* utilising the object for military ends,[228] regardless of "its original nature or…any (later) intended purpose",[229] and regardless of the *extent* of military use.[230] Importantly for a LAWS, this latter point makes it a 'binary concept' in that an attacker need only recognise 'military use' but need not measure its degree or intensity.[231] An example of a situation that may be amenable to machine perception is where enemy forces commandeer civilian cars and taxis,[232] to transport troops/supplies, or merely to use these vehicles as cover. If those persons satisfy the three-/four-part criteria for status-based targeting, their perceptible use of civilian vehicles may make the latter a military objective by 'use', regardless of the extent of that use. Similar reasoning may apply to some other civilian objects like dwellings, a hotel or a school (for troop accommodation, for taking cover, or as observation points)[233], or bridges (for vehicle and troop movements). Insofar as these objects are utilised transparently by enemy combatants, they may become targetable by a LAWS during TLC.

However, in other cases 'military use' can be relatively opaque. For example, power grids and computer hardware and software are unpredictably malleable during an armed conflict, and it is often unclear who is using them.[234] The same can be said of dwellings and other civilian buildings, when used in discreet ways (as a military storage facility via underground tunnels). In yet other cases, the problem is less opacity and more a lack of machine-perceptibleness: consider a civilian broadcast facility used

[228] *AP I Commentary*, ¶ 2022.

[229] Dinstein (n 199), 149.

[230] Schmitt and Widmar (n 144), 393 (emphasis added).

[231] Ibid. (noting that "the object…qualifies as a military objective once it is converted to military use, *however slight*") (emphasis added).

[232] Dinstein (n 199), 149; (n 6), 111 (citing the celebrated 'Taxis of the Marne', commandeered in September 1914 to transport French reserve troops to the frontline, thereby saving Paris from the advancing German forces).

[233] *AP I Commentary*, ¶ 2022; *AMW Manual Commentary*, Rule 22(d), ¶ 2.

[234] Sassóli (n 192), 7.

for military transmission and enemy propaganda.[235] LAWS algorithms will be trained in advance and in relatively abstract settings, yet it is difficult – if not impossible, in the case of computers – to identify when, how and in which context these objects are destined for military use.[236] Similar difficulty will bedevil a LAWS in assessing when discreet (or machine-imperceptible) military use comes to an end, at which point the object ceases to be a lawful target and may no longer be attacked.[237] Accordingly, the pliable concept of *use* at *any given moment* – already very challenging for human combatants to apply – will be nigh-on impossible for a LAWS to assess in the midst of TLC, and in situations of opacity and imperceptibleness.[238]

'Purpose' takes this difficulty to the next level, as it refers to the intended *future* use of an object.[239] Accordingly, 'purpose' is determined *after* the crystallisation of the original 'nature' of an object, but *before* its actual 'use'.[240] This permits the targeting of a civilian object in between uses, and even prior to initial use,[241] thus recognising that an attacker need not wait for a civilian object to actually be utilised for military ends before striking it.[242] Tempering this, however, is a requirement that there be a 'reasonable belief' of *actual* intended future use, not just the mere 'potential' or 'objective possibility' for it.[243] As Dinstein asserts:

> Purpose is predicated on intentions known to guide the adversary, and not on those figured out hypothetically in contingency plans based on a 'worst case scenario'.[244]

---

[235] Office of the Prosecutor, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia* (2000) 39 ILM 1257, ¶ 71 *et seq*.

[236] Sassóli (n 192), 7.

[237] Boothby (n 31), 104; *AMW Manual Commentary*, Rule 22(d), ¶ 4. Furthermore, this complicates the proportionality assessment, as when such an object is put to military use it forms part of the 'military advantage anticipated', but when not in military use it is part of the expected collateral damage.

[238] However, this does not necessarily mean LAWS will be precluded altogether from engaging such objects, if the 'use' can be determined by intelligence analysts during the targeting process and tasked to a LAWS in a TS.

[239] *AP I Commentary*, ¶ 2022 ("The criterion of 'purpose' is concerned with the intended *future* use of an object, while that of 'use' is concerned with its *present* function") (emphasis added). See also *AMW Manual Commentary*, Rule 22(c), ¶ 1.

[240] Dinstein (n 199), 148; (n 6), 113.

[241] Henderson (n 162), 59.

[242] *AMW Manual Commentary*, Rule 22(c), ¶ 1; Schmitt and Widmar (n 144), 393.

[243] Henderson (n 162), 60-61; Boothby (n 31), 103.

[244] Dinstein (n 199), 148; (n 6), 114.

At the very least, determining the enemy's future intention requires knowledge of its Tactics, Techniques and Procedures, and the gathering and analysis of intelligence.[245] Even then, reaching a firm and reliable conclusion is not always easy. Sometimes, enemy intentions are "crisply clear", as in the case of overtly-announced plans;[246] other times, intentions are "not so easy to decipher", and will require relatively more painstaking intelligence efforts in advance.[247] This latter scenario entails the assembly of fragmented pieces of information, often of varying degrees of reliability and with no coherent picture.[248] Hence, there is a need to assess a) the reliability of intelligence;[249] and b) either what is missing and where to obtain it, or "conjecture to fill in the missing pieces of the puzzle".[250] To add further to the cognitive task, conjecture itself must remain consistent with the 'reasonable belief' standard.[251]

This all leads to two pertinent conclusions. First, the determination of a military objective by purpose clearly calls for autonoetically metacognitive thinking,[252] which is a uniquely human domain.[253] This is underscored by the *AMW Manual Commentary*, which advises that:

> The attacker must always act reasonably…[and] ask itself whether it would be reasonable to conclude that the intelligence [regarding future intentions] was reliable enough to conduct the attack in light of the circumstances ruling at the time.[254]

As discussed in 3.2.2.2.1, applying such broad standards as 'reasonableness' to concrete facts, along with the degree of introspection implied here, involves higher-order thinking skills that will arguably not be automated in the near-term.

---

[245] Ford (n 73), 440; subject to foreseeable actions in context, n 246.

[246] Dinstein (n 199), 148; (n 6), 114 (also noting, at 126, foreseeable actions in particular contexts, such as urban warfare waged house-to-house, where the whole block of dwellings may become a military objective by purpose).

[247] Ibid.

[248] Ibid.; *AMW Manual Commentary*, Rule 22(c), ¶ 3

[249] *AMW Manual Commentary*, ibid.

[250] Dinstein (n 6), 114.

[251] Boothby (n 31), 104.

[252] See 3.2.2.2, especially the metacognitive skills and traits listed at page 96.

[253] Except for certain narrowly-defined situations that may be amenable to machine stochastic reasoning, such as the urban warfare scenario suggested by Dinstein at n 246.

[254] *AMW Manual Commentary*, Rule 22(c), ¶ 3.

The second conclusion is that the above account of intelligence activities to establish 'purpose' would seem to describe the kinds of tasks that occur in the deliberate targeting process; in particular, during Phase 2 (target development).[255] Accordingly, the difficulty of establishing enemy intentions and the 'purpose' of an object does not necessarily preclude a LAWS from engaging such objects. Indeed, human pilots currently do not attempt to establish the 'purpose' of an object, but instead operate their planes and weapon systems to complete missions in accordance with the 'target package' provided to them.[256] The higher-order thinking therefore takes place during the earlier phases of the largely human-controlled targeting cycle, with the human pilot merely executing a TS. Even in a dynamic targeting scenario, pilots are briefed with necessary information to engage a target that has, nonetheless, been subjected to considerable human analysis and pre-selected by other specialist personnel.[257] Arguably, there is no reason to believe that a LAWS cannot engage a military objective by purpose in the same way.

'Use' and 'purpose' both make clear that LOAC incorporates a dynamic element, as civilian objects are liable to become military targets depending on the plans, actions and behaviours of both parties.[258] Unlike the analysis of large military objectives by nature or location, both of which rely more on *quantitative* matching (automatic processing), the legitimacy of attacking much of the above is highly fluid and context-dependent. This points to sophisticated *qualitative* analysis (controlled processing) to which humans are predisposed,[259] and this situation is likely to remain true for the foreseeable future.

---

[255] See 5.3.1.2.

[256] See the vignette of an F-16 pilot's mission in Merel Ekelhof, 'Autonomous Weapons: Operationalizing Meaningful Human Control', *ICRC Humanitarian LAW & Policy* (15 August 2018) <http://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/> accessed 21 August 2018 (explaining that the target package typically includes: a description of the target; target coordinates; collateral damage estimates; recommendation of quantity, type and mix of firepower; desired point of impact, to identify aim points; and the weather forecast. Namely, it includes the outputs of Phases 1-5a of the targeting cycle).

[257] Ibid.

[258] Markus Wagner, 'Autonomy in the Battlespace: Independently Operating Weapon Systems and the Law of Armed Conflict' in Dan Saxon (ed.), *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff, 2013), 112.

[259] Wagner (n 145), 1393.

However, contrary to assertions that this makes LAWS unlawful, risky or of limited military value,[260] it merely requires that military objectives by use and purpose be engaged in a TS; or, at most, that LAWS ROE be restricted to military objectives by nature and location. Arguably, these restrictions are not too onerous because, in practice, most attacks are based on an object's nature (in TLC) or use (in a TS),[261] thereby accommodating LAWS deployments.

6.5.3.3 The 'Definite Military Advantage in the Circumstances Ruling at the Time'

Once the first (ECMA) criterion is satisfied, another difficulty may still present itself in the second criterion to be assessed: that the destruction of an object must offer (to the attacker) a 'definite military advantage' (DMA) in 'the circumstances ruling at the time'.[262]

This second prong requires that the military advantage to the attacking party be *definite*, not merely "potential or indeterminate";[263] lest an excessive range of objects become open to attack.[264] Yet, the DMA need not *directly* flow from the attack;[265] nor must it offer immediate tactical gain, but it can be an "operational advantage accruing to the larger campaign".[266]

On the one hand, some commentators consider that this merely duplicates the first criterion,[267] such that the two "mostly presuppose each other".[268] If not always, then at least "most objects" will fulfil both prongs "[a]s a practical matter".[269] If so, a LAWS that is able to meet the first criterion will likely satisfy the principle of distinction *vis-à-vis* objects.

---

[260] Ibid.; Wagner (n 258), 112.

[261] *AMW Manual Commentary*, Rule 22, ¶ 2.

[262] Article 52(2), AP I; *AMW Manual*, Rule 1(y); CIHL, Rule 8.

[263] *AP I Commentary*, ¶ 2024.

[264] Henderson (n 162), 63.

[265] Ibid., 53; Bothe, Partsche and Solf (n 226), 324-25.

[266] Schmitt and Widmar (n 144), 392.

[267] Dinstein (n 6), 104 (arguing destruction of an object that makes an ECMA for one side, will almost always confer a DMA to the other side).

[268] Dill (n 50), 71 (further arguing, at author's n 18, that "the most likely reason why an attack on an object should be militarily advantageous is that it contributes to enemy military action").

[269] Schmitt and Widmar (n 144), 392. See also *AMW Manual Commentary*, Rule 1(y), ¶ 3.

On the other hand, the *AP I Commentary*[270] and several academic commentators[271] take a different view, focusing on the temporal aspect of the second criterion.[272] If this interpretation is accepted, it may require that tactical LAWS units be fed constant updates from the commander on the circumstances of the military operation and its evolution:[273] a requirement that at first blush might seem too onerous, perhaps even undermining the purpose of weapons autonomy. That said, such temporal distinctions are generally rare and occur more as an aberration. In addition, it is arguable that all military objectives by nature can be assumed, by default, to also be military objectives by purpose, so long as they are not completely battle-damaged. This is because of the risk that abandoned military objects may be reoccupied by the enemy and put back to military use,[274] or (if partially damaged) utilised for spare parts. Thus, the second (temporal) criterion may be less relevant to such objects, enabling a LAWS to engage them without needing to undertake complex value judgments.

6.5.3.4 Civilian Objects and Specifically Protected Objects

Once the criteria for military objectives are delineated and applied in a LAWS context, civilian protection becomes easier. Like the analogous provision for persons, Article 52(1), AP I, prohibits making civilian objects "the *object* of attack or of *reprisals*".[275] Again, the nature and wording of this prohibition renders compliance relatively simple: a matter of *ex ante* programming and appropriate deployment,[276] which should pose no difficulty for commanders acting in good faith.[277] Subsequent AP I rules protect specific objects the destruction of which would have an indirectly detrimental effect on civilians. In relation to these, some authors have commented on the limits to

---

[270] *AP I Commentary*, ¶ 2018 (stressing there can only be a 'military objective' in the sense of AP I when the two criteria are "simultaneously present", thereby implying that they can disintegrate).

[271] For example, Boothby (n 31), 103; Henderson (n 162), 48, 51, 55 and 76-78; Hampson (n 51), 49; Timothy LH. McCormack and Helen Durham, 'Aerial Bombardment of Civilians: The Current International Legal Framework' in Yuki Tanaka and Marilyn B. Young (eds.), *Bombing Civilians: A Twentieth-Century History* (The New Press, 2009), 222-24.

[272] McCormack and Durham, ibid., 223 (explaining that in the first Gulf War, coalition forces chose not to attack Iraqi fighter planes, partly because the jets were parked on desert sand and were a significant distance from any airstrip. Thus, the 'circumstances ruling at the time' would have made the attack of questionable military utility).

[273] Nathalie Weizmann, 'Autonomous Weapon Systems under International Law', *Academy Briefing No. 8* (Geneva Academy of International Humanitarian Law and Human Rights, November 2014), 14.

[274] Abandonment and later reoccupation of military vehicles was one of the observation of US forces in Iraq. See Joint Readiness Training Center, 'Operation OUTREACH: Tactics, Techniques, and Procedures', *News Letter No. 03-27* (October 2003).

[275] Article 52(1), AP I (emphasis added).

[276] Schmitt (n 63).

[277] *AP I Commentary*, ¶ 2198.

machine perception and have queried how LAWS will respect these rules.[278] Yet, on closer examination, these prohibitions can also (largely) be seen as programming and deployment matters, which may be expected to pose little or no difficulty for commanders utilising the Joint Targeting process and acting in good faith.

6.5.3.4.1 Cultural Property

For example, Article 53(a), AP I, prohibits "acts of hostility *directed* against…historic monuments, works of art or places of worship".[279] The adverb 'directed' clearly goes to deliberate human choices made during the targeting cycle, and the same can be said about Paragraph (c), which prohibits "mak[ing] such objects the *object* of *reprisals*".[280]

However, the ICRC's restatement of customary law goes further than AP I, and states that:

> Special care must be taken in military operations to avoid damage to buildings dedicated to religion, art, science, education or charitable purposes and historic monuments unless they are military objectives.[281]

In a LAWS context, much of this 'special care' will begin at Phase 2 of the deliberate targeting cycle. For example, using the UNESCO World Heritage List[282] and World Heritage in Danger List,[283] targeteers have an immediate and authoritative basis to enter high-priority sites on the no-strike list, which a LAWS would 'respect' by avoiding any attacks on the relevant GPS coordinates (immovable cultural property) and/or image matches (movable or immovable). Of course, not all cultural sites benefit from a UNESCO listing, so attacking forces may also have to consult other lists. In many cases, however, the most comprehensive and relevant lists of protected heritage (and other protected buildings and monuments) are in the hands of the host State, which has no specific obligation to provide that information to its adversary.[284] On the

---

[278] For example, Wagner (n 145); (n 258); Ozlem Ulgen, 'Definition and Regulation of LAWS' *Submission to April 2018 GGE* (5 April 2018), ¶ 11 <https://www.researchgate.net/publication/324227191_Dr_Ulgen_UN_GGE_LAWS_April_2018_-_submission_-_Definition_and_Regulation_of_LAWS> accessed 21 August 2018.

[279] Article 53(a), AP I (emphasis added); *AMW Manual Commentary*, Rule 95(a).

[280] Article 53(c), AP I (emphasis added).

[281] CIHL, Rule 38(A).

[282] 'World Heritage List' <https://whc.unesco.org/en/list/> accessed 21 August 2018.

[283] 'List of World Heritage in Danger' <https://whc.unesco.org/en/danger/> accessed 21 August 2018.

[284] Marina Lostal, Kristin Hausler and Pascal Bongard, 'Armed Non-State Actors and Cultural Heritage in Armed Conflict' (2017) 24 International Journal of Cultural Property 407, 419-20 ("While all parties

other hand, there may be a general obligation to do this under Article 58, AP I,[285] if not under the Article 1(1), AP I, obligation to respect and to *ensure respect* for LOAC *erga omnes*.[286]

Perhaps a better option, which is *relatively* within the control of attacking forces, is to work with archaeologists to identify all relevant sites that merit protection[287] and to begin this process even before commencement of the formal targeting process, if willing experts can be found.[288] Separately, where a protected object is characterised by distinctive architecture, this may be amenable to the object recognition of an ATR, thus avoidable even in the absence of any list.[289]

Where cultural property becomes a military objective,[290] the *AMW Manual Commentary* advises that the decision to attack be taken by an "appropriate level of command", which is taken to mean at least an air squadron or battalion commander.[291] Further, such a decision is to be made "with due consideration of its special character as cultural property", as such decisions "cannot be taken lightly".[292] This clearly involves complex value judgments, which implicate human metacognitive thinking and discretion; again going back to the human-led targeting process in deploying LAWS for a TS. Namely, attacks on cultural property cannot be lawfully executed through generalised parameters programmed for TLC, as the automatic processing of the control software will not be able to make the necessary value judgments.

---

to the conflict must respect cultural property, knowing its location is a prerequisite for this to happen. However, the Hague Convention does not envisage the exchange of relevant information between warring parties").

[285] See 7.3.3.

[286] As also seen in the customary rule restated in CIHL, Rule 144.

[287] See, for example, Peter Stone, 'The Identification and Protection of Cultural Heritage During the Iraq Conflict: A Peculiarly English Tale' (2005) 79 Antiquity 933 (outlining the author's role in advising the UK Ministry of Defence on cultural sites to be protected in Iraq, both before and during the 2003 invasion).

[288] This is not always easy, as some archaeologists with the right expertise may be opposed to assisting military preparations. See, for example, John Curtis, 'Relations Between Archaeologists and the Military in the Case of Iraq' in Peter G. Stone (ed.), *Cultural Heritage, Ethics and the Military* (Boydell Press, 2011). More broadly, archaeologists face a plethora of complex considerations in deciding whether to assist, thus their participation is not guaranteed. See Umberto Albarella, 'Archaeologists in Conflict: Empathizing with Which Victims?' (2009) 2 Heritage Management 105.

[289] For example, distinctive church or mosque architecture, which an ATR can be trained to recognise via supervised learning; on which, see 2.5.1.3.

[290] As anticipated in CIHL, Rule 38(B); and *AMW Manual Commentary*, Rule 95(b), ¶ 4. Presumably this would occur by location, use or purpose only.

[291] *AMW Manual Commentary*, Rule 96, ¶ 5.

[292] Ibid., Rule 96, ¶ 6.

6.5.3.4.2 Objects Indispensable for Civilian Survival

Article 54(2) prohibits attacks against "objects indispensable to the survival of the civilian population…for the *specific purpose* of denying them for their sustenance value".[293] Once again, the wording of the provision clearly indicates human choices made through the targeting cycle. This is underscored by the UK's and France's statements of interpretation upon ratifying AP I that Article 54(2) does not apply to attacks carried out for a specific purpose *other than* denying sustenance to the civilian population.[294] Moreover, the *AMW Manual Commentary* emphasises the need for a 'specific purpose' and precludes "incidental distress of civilians resulting from otherwise lawful military operations".[295] Accordingly, so long as commanders – supported by multiple battle staffs and legal advisers, and overseen by the Joint Targeting Coordination Board – do not deploy LAWS to attack such indispensable objects *for the specific purpose* of denying sustenance to the civilian population, or as a reprisal,[296] compliance with Article 54(2) should be relatively easy.

6.5.3.4.3 Infrastructure That May Release Dangerous Forces

Article 56(1) prohibits making the object of attack "[w]orks or installations…[that] may cause the release of dangerous forces and consequent severe losses among the civilian population", even if such works or installations are military objectives.[297] As with the last two prohibitions, this one also does not pose insurmountable compliance difficulty for a LAWS-deploying Belligerent, which utilises a formal targeting process. First, it is significant that the protection from direct attack is limited to three specific types of infrastructure: dams, dykes and nuclear electrical generating stations.[298] Together with the cumulative nature of the criteria[299] and the focus on *ex*

---

[293] Article 54(2), AP I; *AMW Manual*, Rule 97(b) (providing examples of foodstuffs, agricultural areas for the production of foodstuffs, crops, livestock, drinking water installations and supplies, and irrigation works) (emphasis added). See also the customary rule restated in CIHL, Rule 54 (though excluding the 'specific purpose' qualifier).

[294] UK, Reservations and Declarations Made Upon Ratification of AP I (28 January 1998), Statement (l); France, Reservations and Declarations Made Upon Ratification of AP I (11 April 2001), ¶ 14.

[295] *AMW Manual Commentary*, Rule 97(b), ¶ 2.

[296] Paragraph (4) of Article 54 prohibits making such objects "the object of reprisals".

[297] Article 56(1), AP I. See also the restatements of this rule in *AMW Manual*, Rule 36; CIHL, Rule 42 (both applying a lower standard).

[298] *AP I Commentary*, ¶¶ 2147-2150.

[299] Leslie C. Green, *The Contemporary Law of Armed Conflict* (2nd ed., MUP, 2000), 158 (explaining that the two criteria are cumulative ("and consequent"), such that an attack causing the releases of dangerous forces *away* from an urban centre, thus not liable to cause 'severe' civilian losses, would potentially be lawful).

ante scrutiny,[300] this limits the burden on intelligence analysts during Phase 2 of the deliberate targeting cycle (or the *target* stage of the dynamic cycle), and it should facilitate the compilation of a definitive list of such objects and their precise locations. Second, the protection is "unique" in that it continues even when the works or installations are put to military use and thus "glaringly constitute military objectives".[301] Arguably, the combined effect of these two factors is to create an administrable no-strike category: a set of a 'binary actions' that are amenable to both pre-deployment programming and in-field machine perception via GPS guidance systems.[302]

Furthermore, given the possibility to integrate collateral damage estimation capabilities into LAWS,[303] this argument may even extend to the second prohibition in Article 56(1), against attacking *other* nearby military objectives, if such an attack may also cause the release of dangerous forces and consequent severe civilian losses. Thus, underlying the prohibition is a "worst case analysis", which assumes that such attacks will induce "massive risks" to the civilian population. Specifically, these risks are assumed to be a) "unacceptably high", b) almost never outweighed by military advantage and, thus, c) cannot be justified by any claim of military necessity, except under the three specific exceptions in Article 56(2).[304] Again, the specificity of the rule may be expected to support machine application

The fact that protection under Article 56(1) is qualified by the verb 'may' and the adjective 'severe' does not imply that a LAWS will have to undertake any value judgments. Rather, as the *AP I Commentary* points out, 'severe' (losses among the civilian population) is a matter of 'common sense' and is to be applied in 'good faith' on the basis of objective criteria, such as population density and the proximity of inhabited areas.[305] Accordingly, commanders and their battle staffs are to make these

[300] Frits Kalshoven, *Reflections on the Law of War: Collected Essays* (Martinus Nijhoff, 2007), 235 (explaining that the commander's judgment is assessed in light of all information available at the time the attack was planned or launched, and whether these objectively pointed to the possibility of severe losses to the civilian population).

[301] Dinstein (n 6), 227.

[302] Boothby (n 81), 81.

[303] See 7.2.1.

[304] Oeter (n 187), 218.

[305] *AP I Commentary*, ¶ 2154.

judgment calls when deciding which specific objects and locations to put on the no-strike list, which a LAWS will simply be programmed not to attack.

Arguably, the same is true with respect to the specific grounds on which protection from attack shall cease under Article 56(2). These require that a) the work, installation or nearby military objective is used in *regular*, *significant* and *direct support* of military operations, and b) that an attack is the only feasible way to terminate such support. This sets the bar significantly higher than the 'effective contribution to military action' that an object must make to qualify as a military objective under Article 52(2) and it calls for a commander "at the highest military level" to make the judgment call,[306] usually based on prior intelligence. This again points to the deliberate (or at least the dynamic) targeting cycle in reaching a deliberative human decision to conduct an attack pursuant to Article 56(2), while a 'narrow loop' LAWS will merely execute the attack via TS, and will *refrain* from such actions at all other times.

6.5.3.4.4 Medical Capabilities

The protection of medical capabilities to treat the sick, wounded and shipwrecked is a particular concern in IHL/LOAC, and is an essential component of efforts to humanise war. To this end, there are specific distinction-based AP I rules that afford respect for, and protection to:

- Fixed and mobile medical units,[307] with Parties to the conflict being encouraged to notify each other of the locations of fixed units.[308]
- Medical vehicles used exclusively for transportation.[309]
- Hospital ships and coastal rescue craft.[310]
- 'Other' medical ships and craft.[311]
- Medical aircraft.[312]

Like the previous prohibitions, these are also simple programming and deployment matters. However, as the *AMW Manual Commentary* points out, "to respect" medical

[306] Ibid., ¶ 2159

[307] Article 12(1), AP I; Article 19, GC I; Article 18, GC IV; CIHL, Rule 28.

[308] Article 12(3), AP I; *AMW Manual*, Rule 73.

[309] Article 21, AP I; Article 35, GC I; Article 38, GC II; Article 21, GC IV; CIHL, Rule 29.

[310] Article 22, AP I; Articles 22, 24, 25, 27 and 28, GC II; CIHL, Rule 28.

[311] Article 23, AP I, also expanding on analogous provisions in GC II.

[312] Article 24, AP I; Article 36, GC I; Article 39, GC II; CIHL, Rule 28. *AMW Manual*, Rule 72(a).

personnel and facilities is broader than simply refraining from directly attacking them, and it includes a prohibition against "unnecessarily preventing them from discharging their functions".[313] Thus, a LAWS would have to be programmed to, for example, keep a distance from such facilities, lest it inadvertently creates a fear of impending attack, thereby disrupting medical operations. Crucially, recognition of protected status – an essential prerequisite for respecting it – can be greatly enhanced via technical means, which exploit the strengths of automatic processing.

6.5.3.4.5 Enhancing Detection by Technical Means

First, as noted above, whenever the location of a fixed protected object is known, respect will be effectuated primarily by assigning a no-strike categorisation to its GPS coordinates.[314] Beyond this, and in the case of unknown or movable objects, there are additional safeguards, which may support autonomous attack and make it more discriminating. These include:

- The Blue Shield that denotes cultural property.[315]
- The international special sign (three bright orange circles) for works and installations containing dangerous forces.[316]
- The distinctive emblems of the Red Cross and Red Crescent,[317] which denote medical and religious personnel and facilities.[318]

Importantly, these can all be specifically designed to facilitate detection by the ATR of a LAWS; for example, using ancillary lighting, thermal ribbons and detailed colour contrasts.[319]

---

[313] *AMW Manual Commentary*, Rule 71, ¶ 12 (giving the example of blocking medical supplies).

[314] Boothby (n 81), 81. In addition, GPS coordinates will aid respect for non-defended localities and demilitarised zones under Articles 59 and 60, AP I (restated in CIHL, Rules 37 and 36), respectively. The terms, conditions and boundaries of both are extensively deliberated by human negotiators, after which LAWS need only be programmed correctly to comply with those (spatio-temporal) boundaries.

[315] Articles 6, 10, 16, 17 and 20, Convention for the Protection of Cultural Property in the Event of Armed Conflict (adopted 14 May 1954, entered into force 7 August 1954) 249 UNTS 240.

[316] Article 56(7), AP I; Article 17, Amended Annex I.

[317] Article 4, Amended Annex I; CIHL, Rule 30.

[318] Article 18, AP I, provides that each Party "shall endeavour" to ensure that relevant personnel and facilities are identifiable; and shall implement methods and procedures to enhance recognition, mainly via the distinctive emblems, but also using the 'distinctive signals' in addition or in lieu (on which, see below).

[319] Article 17(4), Amended Annex I, states that in times of reduced visibility, "the [international special] sign may be lighted or illuminated…[and] made of materials rendering it recognizable by technical means of detection". Article 5(3) provides the same in relation to the distinctive emblem, and it suggests a method of painting the emblem that should "facilitate its identification, in particular by infrared

Furthermore, there is a range of 'distinctive signals' for the exclusive use of medical units and transports; for example, the distinctive light signal,[320] radio signals and radio messages,[321] and various forms of electronic identification.[322] Each is individually predisposed to relatively reliable detection by technical means, and in combination with each other and with the distinctive emblems, they offer an invaluable means of detection-confirmation. These will further enhance the distinction capabilities of a LAWS, and help to avoid the kinds of unintended engagements seen with manned targeting.[323]

That said, attacking forces must remain vigilant and avoid any over-reliance on emblems, signs and signals. Indeed, placing too much faith in these safeguards – and in human efforts to deploy them fully and accurately – may lead to a watering down of commander-led targeting efforts and, ultimately an increase in distinction failure.

6.5.3.4.6 Enhancing Confidence in Technical Detection

Helping to enhance the confidence of attacking forces in their ATR assessments, Article 38, AP I, prohibits adverse Parties from making any improper use of emblems, signs or signals;[324] for example, by attaching them to military objectives. Moreover,

---

instruments". This is significant for autonomous attack, as infrared instruments will be a standard feature on LAWS. See also *AMW Manual Commentary*, Rule 72(b) and its accompanying commentary.

[320] Article 7, Amended Annex I, recommends a distinctive blue light signal with specific boundaries for its chromaticity and flashing rate. This will clearly aid recognition by the visual and light sensors of a LAWS.

[321] Article 8, Amended Annex I, provides that the radio signal shall consist of a standardised urgency signal and distinctive signal; and that any subsequent radio message is transmitted in English and will convey specific data, such as the call sign of the medical transport, its position, intended route, etc. This will aid detection by a LAWS fitted with a radio receiver and speech recognition software.

[322] Article 9, Amended Annex I, provides that radar transponders and underwater acoustic signals may be used to identify protected medical aircraft and vessels, via standardised/agreed upon codes. This will clearly aid detection by autonomous drones installed with the Secondary Surveillance Radar system, or autonomous undersea vehicles installed with, for example, a hydrophone.

[323] See the numerous incidents of wrongful targeting, which were largely down to human error in competences involving automatic processing. For example, where war-weary pilots have failed to check the no-strike list and have not paid sufficient attention to clearly visible protective signs and emblems, this has resulted in erroneous attacks on hospitals and ICRC facilities. The result has often been a reduced capacity to distribute humanitarian supplies and medical services on the ground. Yet, autonomous drones operating with sophisticated ATR and GPS guidance systems would almost certainly have avoided those attacks. See 'ICRC Warehouses Bombed in Kabul', *ICRC News Release 01/43* (16 October 2001) <https://www.icrc.org/eng/resources/documents/news-release/2009-and-earlier/57jrcz.htm>; 'Bombing and Occupation of ICRC Facilities in Afghanistan', *ICRC News Release 01/48* (26 October 2001) <https://www.icrc.org/eng/resources/documents/news-release/2009-and-earlier/57jrdx.htm>; 'Kunduz Bombing: US Attacked MSF Clinic 'In Error'', *BBC News* (25 November 2015) <http://www.bbc.co.uk/news/world-asia-34925237>; all accessed 21 August 2018.

[324] Article 38, AP I; *AMW Manual*, Rule 112(a)-(b); CIHL, Rules 59-61.

should such improper use become perfidious,[325] this will elevate the violation to a gross breach of AP I and, therefore, a war crime.[326] The discussion on perfidy in 6.5.2.1.2 applies equally here, including the likelihood that adversarial examples, which merely imitate the recognised emblems, signs or signals, will be sufficient to establish a violation.

6.5.4 Will LAWS be Able to Sense Targeting 'Doubt'?

It was argued in 3.2.2.2.1 that the decision to attack is often based on incomplete or inconclusive information. The resulting uncertainty (or 'fog of war'), which pervades armed conflict raises the question: "how certain must [an attacker] be that the object or person is a lawful target before proceeding?"[327] As a matter of law, both the question and its answer are crucial, as AP I mandates that in the event of 'doubt', civilian status shall be presumed for both persons[328] and objects,[329] thereby protecting them from direct attack.[330] Importantly, this presumption relates to civilian *status* and is not a conduct-based presumption against DPH.[331] In relation to objects, the language of the AP I norm is reproduced in Amended Protocol II,[332] which regulates the use of anti-personnel mines. Significantly, these weapons also operate with humans out-of-the-narrow-loop, thereby underscoring the need to act on doubt in such circumstances.[333] With these in mind, the *AMW Manual Commentary* fully extends the rule of doubt to

[325] See 6.5.2.1.2 on the elements of perfidy and their application to LAWS.

[326] Article 85(3)(f), AP I.

[327] Ian Henderson and Bryan Cavanagh, 'Unmanned Aerial Vehicles (UAVs): Do They Pose Legal Challenges?' in Hitoshi Nasu and Robert McLaughlin (eds.), *New Technologies and the Law of Armed Conflict* (TMC Asser Press, 2014), 204.

[328] Article 50(1), AP I: "In case of doubt whether a person is a civilian, that person shall be considered to be a civilian." See also CIHL, Rule 6.

[329] Article 52(3), AP I: "In case of doubt whether an object which is normally dedicated to civilian purposes…is being used to make an effective contribution to military action, it shall be presumed not to be so used." See also CIHL, Rule 10.

[330] And requiring that they be taken into account in proportionality and precautionary considerations.

[331] Boothby (n 31), 149, 427; Cf. ICRC (n 138), 75-76. Recall from 6.5.2.3 that near-term LAWS are, in any event, unlikely to be given target engagement authority in DPH situations, because of the high metacognitive demands.

[332] Article 3(8)(a), Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices (adopted 10 October 1980, amended 3 May 1996, entered into force 3 December 1998) 2048 UNTS 93.

[333] Though it is acknowledged that the doubt is addressed to the human emplacing the mine, not the mine itself. Even so, it highlights the importance of an analogous exercise of doubt in a LAWS context, when planning a TS.

autonomous lethal targeting,[334] where it will be a prominent factor in both the development and deployment of LAWS.[335]

That said, the *degree* of doubt required to trigger the presumption of civilian status is not codified in treaty law, and with varying State practice there is arguably no customary standard.[336] To be sure, war is replete with uncertainty, and the mere existence of *some* doubt is insufficient to preclude an attack;[337] rather, as the *AP I Commentary* makes clear, 'doubt' is likely to be context-specific.[338] Accordingly, the ICTY Trial Chamber in *Galić* articulated the legal standard as:

> *…when it is not reasonable to believe*, in the *circumstances of the person contemplating the attack*, including the *information available* to the latter, that the potential target is a combatant [or an object being used to make an effective contribution to military action].[339]

The *AMW Manual Commentary* echoes this,[340] as does the 'positive identification' (PID) standard set out in some ROE, which requires "a *reasonable certainty* that the proposed target is a legitimate military target".[341] Henderson and Cavanagh therefore argue that 'reasonable belief' and 'reasonable certainty' are practically synonymous and, when considered in the circumstances of the attacker – including the information or intelligence available to him – provide a sufficiently clear and practical test;[342] at least for a metacognitive human.

---

[334] *AMW Manual Commentary*, Rule 39, ¶ 5 ("The standards…regarding doubt apply equally to UCAV attacks, *whether autonomous or manned*") (emphasis added).

[335] Thurnher (n 59), 191.

[336] Michael N. Schmitt (ed.), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd ed., CUP, 2017) (hereafter, *Tallinn Manual 2.0*), Rule 95, ¶ 3 (noting that, despite no definitive boundary, the mere existence of *some* doubt is insufficient).

[337] *Tallinn Manual 2.0*, Rule 95, ¶ 3 (persons) and Rule 102, ¶ 9 (objects); *AMW Manual Commentary*, Rule 12(a), ¶ 4 (persons) and Rule 12(b), ¶¶ 4 and 5 (objects).

[338] *AP I Commentary*, ¶ 1920 (referring to those "whose status seems doubtful because of the circumstances").

[339] *Prosecutor* v. *Galić* (ICTY Trial Judgment) IT-98-29-T (5 December 2003), ¶ 50 (persons), ¶ 51 (objects), (emphasis added). See also ¶ 55 ("the Prosecution must show that in the given circumstances a reasonable person could not have believed that the individual he or she attacked was a combatant").

[340] *AMW Manual Commentary*, Rule 12(a), ¶ 4 ("The degree of doubt necessary to preclude an attack is that which would cause a reasonable attacker in the same or similar circumstances to abstain from ordering or executing an attack").

[341] See CFLCC and MNC-I ROE Cards, reprinted in LCDR David H. Lee (ed.), *Operational Law Handbook* (JAG's Legal Center & School, US Army, 2015), 109-10 (emphasis added).

[342] Henderson and Cavanagh (n 327), 205.

In a LAWS context, this means where there is enough doubt that a *reasonable human attacker* – possessing the *same information* and in *a similar situation as the LAWS* – would hesitate, then an attack will not legally be allowed to proceed.[343] In such a case of uncertainty, the LAWS must be programmed to a) recognise the situation of 'doubt' that would cause a human to hesitate and b) abort the attack,[344] or at least contact a human operator for further instructions.

Yet, for the reasons discussed in 3.2.2.2.1, this framing of doubt in human reasonableness terms will complicate translation into a LAWS context.[345] A significant challenge will be to develop an automated mechanism that a) accurately gauges doubt, and b) reliably factors in the unique situation in which the LAWS is operating.[346] In this regard, the Trial Chamber in *Galić* noted that observations relating to the clothing, activity, age, or gender are relevant when determining whether a person is a civilian[347] and, therefore, whether there is enough doubt to trigger the presumption of civilian status. Other factors have already been mentioned, above, again in relation to persons.[348] However, it is not clear how amenable to automatic processing these will be in any given battlefield. Certainly, in the most dynamic battlefields there will potentially be relevant factors that are not foreseen by programmers (or by case law), but to which metacognitive human combatants would be able to improvise.

It should be noted that in relation to objects, the rule is not about doubt in general, but specifically about whether a civilian object is being put to military 'use'.[349] In that regard, the *Tallinn Manual 2.0* points out that in establishing doubt *versus* the reasonableness of an assessment of military use, an attacker should consider:

> [T]he apparent reliability of the information, including the credibility of the source or sensor, the timeliness of the information, the likelihood of deception, and the possibility of misinterpretation of data.[350]

---

[343] Thurnher (n 335), 191-92.

[344] Wagner (n 258), 113; Weizmann (n 273), 14.

[345] Schmitt (n 63), 16-17

[346] Ibid.; Thurnher (n 335), 192.

[347] *Prosecutor* v. *Galić* (Trial), ¶ 50.

[348] See (notes and text accompanying) nn 146-147.

[349] See n 329. Thus, the rule of doubt does not extend to military objectives by nature, location or purpose.

[350] *Tallinn Manual 2.0*, Rule 102, ¶ 8.

Namely, in case of any doubt as to whether a civilian object is making an ECMA by use, it may only be attacked after a "careful assessment" of the situation.[351] This clearly calls for the marshalling of higher-order metacognitive skills, which a LAWS will not possess. It was one of the reasons argued at 6.5.3.2, for why military objectives by use (or purpose) are likely to be engaged only via a TS. Namely, the need for extensive intelligence analysis and human deliberation will likely demand the rigours of the deliberate (or at least the dynamic) targeting cycle, led by human decision-makers.

On the other hand, it was also argued at 6.5.3.2 that where clearer, more machine-perceptible instances of military 'use' are detected, this may be amenable to autonomous attack. Specifically, in relation to aircraft, the *AMW Manual Commentary* provides an illustrative list of factors, which are potentially relevant to recognising 'doubt' in air warfare.[352] Processing these data is similar to the quantitative matching mentioned in 2.5.4.1, in relation to ATR. In a similar vein, Arkin discusses recognition of 'uncertainty' through a weighted average of discrete values, e.g. binary (absent or present), or categorical (absent, weak, medium, strong); or real continuous values.[353] This may also combine with 'conservative use of lethal force', where a LAWS – not affected by any survival instinct – can hold fire to resolve doubt.

Ultimately, whether a LAWS can administer targeting doubt to the legally required standard will – like much of the above analysis – depend on the system, the task being programmed and the operational environment. The more complex and dynamic the task and environment, the more likely the system will be legally non-compliant for lack of controlled processing. Conversely, the simpler and more static the task and environment, the more likely targeting doubt can be resolved through overlapping criteria and statistical confidence thresholds (i.e. automatic processing). Perhaps the more important 'doubt' that needs to be taken into account is that of the commander, when deploying systems into specific missions.

---

[351] Ibid. See also CIHL, Rule 10 (stating "in case of doubt, a careful assessment has to be made under the conditions and restraints governing a particular situation as to whether there are sufficient indications to warrant an attack).

[352] *AMW Manual Commentary*, Rule 40, ¶ 4 (a)-(i) (listing visual image; response to warnings; infrared, radar and electromagnetic signatures; identification modes and codes; number and formation of aircraft; altitude, speed, track, profile and other flight characteristics; pre-flight and in-flight air traffic information).

[353] Arkin (n 154), 59.

6.5.5 Ensuring the Compliance of LAWS Operations with the Principle of Distinction

On balance, therefore, compliance with the principle of distinction is not just contingent on system capabilities, but is more about commanders undertaking an early 'matching exercise'. That is, to take into account the assigned task and the prevailing operational environment; to understand system capabilities, limitations and how the LAWS will interact with its task and environment; and to ensure that one is appropriately matched to the other.[354]

Thurnher elaborates on this by pointing to the spectrum of battlefields, from the most uncluttered right through to full urban warfare.[355] Clearly, as the task and environment become more complex, the demands on the LAWS sensory capabilities will increase, with progressively more robust sensor packages and ATR software becoming necessary. The author concludes:

> Ultimately, autonomous weapons will only be an option in areas where the systems are able to reasonably distinguish between combatants and civilians and between military objectives and civilian objects given the particular battlefield circumstances ruling at the time.[356]

A similar position is taken by Anderson, Reisner and Waxman[357] and Boothby,[358] while Crootof[359] and Scharre[360] point out that in simple environments, a degree of autonomous operation is already achievable. In this connexion, the April 2018 Group of Governmental Experts (GGE) meeting saw a canonical example of a LAWS currently in development being discussed:

---

[354] Schmitt (n 63). Note that this is also the basis on which the landmines regime works: to prohibit deployment where distinction is not possible, due to a cluttered operational environment and/or limitations of the munitions' capabilities.

[355] Thurnher (n 335), 188.

[356] Ibid.

[357] Kenneth Anderson, Daniel Reisner and Matthew C. Waxman, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems' (2014) 90 International Law Studies 386, 401-02 (giving the canonical examples of an aerial dogfight over the high-seas and undersea attacks on submarines).

[358] Boothby (n 81), 78-79 (also emphasising the importance of the weapons review process in demarcating this).

[359] Rebecca Crootof, 'The Killer Robots are Here: Legal and Policy Implications' (2015) 36 Cardozo Law Review 1837, 1874-75.

[360] Scharre (n 164).

> [A]n underwater autonomous vessel equipped with a sonar, ship registry data, and torpedoes [that] would be able to recognise and differentiate between civilian and military vessels based on the input from the sonar system and comparison of the input with the onboard ship registry. In case a civilian vessel is detected, the torpedoes would not be launched or would be diverted.[361]

Being in a simple and structured environment, to pursue cooperative targets using appropriate sensors and data on potential civilian presence, this particular LAWS is unlikely to pose a significant distinction challenge.

However, even in a complex and dynamic battlefield with a high degree of comingling, restricting the ROE of a LAWS may offer another way to comply with the principle of distinction in TLC. Namely, DPH is a permissive exception while military 'use' and 'purpose' are merely alternative targeting criteria: none of these necessarily have to be invoked, much less so in an autonomy context. As Backstrom and Henderson explain:

> [I]f a commander was prepared to forgo some theoretical capability, it is possible…to confine the list of targets that are subject to automatic target recognition to a narrow list of objects that are clearly military objectives by their nature.[362]

The same is potentially true in relation to traditional combatants, who may be identifiable using the three-/four-part criteria. By restricting LAWS ROE in this way, commanders will exploit the speed and precision of automatic processing, while human combatants pursue the more abstract and malleable categories, which require nuanced judgment and controlled processing to determine their battlefield status.[363] Alternatively, those categories may be engaged in a TS, where the controlled processing and extensive metacognitive thinking is undertaken by a plethora of personnel in advance, leaving the LAWS to execute the strike with machine-precision. In all these cases, there is clearly a 'division of labour' between man and machine, based on their respective cognitive strengths.

---

[361] 'Chair's Summary of the Discussion on Agenda item 6(a) 9 and 10 April 2018, Agenda item 6(b) 11 April 2018 and 12 April 2018, Agenda item 6(c) 12 April 2018, Agenda item 6(d) 13 April 2018', *Chair's Documents at the 2018 GGE Meeting on LAWS* (9-13 April 2018), 6-7 <https://www.unog.ch/80256EDD006B8954/(httpAssets)/DF486EE2B556C8A6C125827A00488B9E/$file/Summary+of+the+discussions+during+GGE+on+LAWS+April+2018.pdf> accessed 21 August 2018.

[362] Backstrom and Henderson (n 28), 492.

[363] Ibid.

Complementing Backstrom and Henderson, Canning also proposes restricting LAWS (or at least their autonomous capabilities) to anti-material engagements.[364] Namely, for machines to target machines (mostly weapons), while humans target humans.[365] This is aptly described as targeting "either the 'bow' or the 'arrow', but not the human 'archer'".[366] The idea is to use *non-lethal* force to convince the enemy to abandon its weapons, before destroying the latter and removing the former's destructive capabilities;[367] people may still be killed, but only as a secondary consequence of targeting the weapon.[368] This proposal aims to address the distinction challenge by assuming that military-grade weapons and other military equipment, which are relatively more amenable to ATR, are only of interest to combatants or civilians who undertake DPH. There is some disagreement with this assumption,[369] though it is submitted that any residual targeting errors would be the same, if not worse in the case of manned targeting,[370] and can often be mitigated by civilians exercising due caution.[371] While imperfect on its own, Canning's proposal does suggest another way for LAWS to be deployed in compliance with the principle distinction and, if the author's "dream machine" is realised, may even offer a deployment option in urban warfare.[372]

---

[364] John S. Canning, 'A Concept of Operations for Armed Autonomous Systems', *Presentation at Third Annual Disruptive Technology Conference* (6-7 September 2006), 14-16 <https://ndiastorage.blob.core.usgovcloudapi.net/ndia/2006/disruptive_tech/canning.pdf> accessed 21 August 2018.

[365] Ibid., 14 (suggesting that the same LAWS may be used throughout, but with 'dial-a-level' autonomy to switch from one mode to another).

[366] John S. Canning, 'You've Just Been Disarmed. Have a Nice Day!' (2009) 28 IEEE Technology and Society Magazine 12, 14.

[367] Canning (n 364), 14.

[368] Ibid., 17.

[369] For example, Wagner (n 145), 1391 (arguing that civilians with a legitimate use for rifles – be that for cultural reasons, self-defence or hunting – may still be wrongly targeted).

[370] See, for example, Luke Harding and Matthew Engel, 'US Bomb Blunder Kills 30 at Afghan Wedding', *The Guardian* (2 July 2002) <https://www.theguardian.com/world/2002/jul/02/afghanistan.lukeharding> accessed 21 August 2018 (reporting on one of many wedding party incidents in Afghanistan, where US forces mistook celebratory gunfire as a threat, thereby returning fire and killing 30 wedding guests, all civilians).

[371] Dinstein (n 6), 141 (advising civilians not to brandish self-defence weapons near the contact zone, and to postpone recreational shooting until calmer times).

[372] Canning (n 366), 14 (positing a weapon system that would confront an enemy soldier/civilian DPH, physically remove his rifle and saw it in half with a diamond-tipped saw; leaving an unarmed, yet unharmed individual).

To summarise, the lawful use of LAWS is possible where the system can reasonably distinguish between the relevant categories, given its ATR and the specific battlefield circumstances ruling at the time; and where appropriate constraints and parameters can be put on its engagement options. This makes compliance with the principle of distinction largely a programming and deployment issue rather than one that should be judged on the state of technology alone. Concretely, it speaks largely to the pre-deployment phases of the Joint Targeting Cycles.

6.6 Conclusion

Distinction is undoubtedly one of the most important of LOAC norms, yet it is clearly not easy to comply with in every situation. Yet, by using common sense and acting in good faith, commanders can *in principle* find suitable restrictions and precautions as to deploy LAWS appropriately, and in a way that adequately distinguishes lawful from unlawful targets. Of course, this assumes a) effective training of commanders, and full knowledge on both the capabilities and limitations of prevailing ATR systems, and b) a degree of self-restraint on the part of those commanders, who may be operating under extraordinary operational pressures. Arguably, neither of these conditions will necessarily hold true all of the time, and certainly not on all sides of an armed conflict. Thus, while compliance with the principle of distinction is possible, this will only result from assiduous and well-informed advance decision-making by genuinely accountable commanders.

Chapter 7
Targeting Law II: Can LAWS Be Deployed in Compliance with the Principle of Proportionality and with Adequate Precautions?

7.1 Introduction

The following chapter continues the targeting law analysis of lethal autonomous weapon systems (LAWS) under both treaty[1] and customary[2] international humanitarian law (IHL)/law of armed conflict (LOAC). It will continue bearing in mind the factors bullet-pointed at the start of 6.4, and will continue gleaning insights from the *AP I Commentary*,[3] and the *Air and Missile Warfare Manual*[4] and its *Commentary*,[5] where relevant. The examination begins at 7.2 with the highly amorphous principle of proportionality, first clarifying some pivotal terms in 7.2.1 before exploring the problematic nature of its application in a LAWS context, in 7.2.2. As will be seen here, proportionality compliance will pose one of the biggest challenges for LAWS deployments, because of a) the *shifting* and *contextual* nature of military advantage, b) the *incommensurability* between that and collateral damage, and c) the *indeterminate* nature of 'excessiveness'. Collectively, these call for highly deliberative thinking, which would tend to negate the automated application of proportionality. That said, there are potential ways to guard against excessive collateral damage by a LAWS, which are briefly considered at 7.2.3. These are mainly rooted in the targeting process and the fact that collateral damage estimation is highly amenable to automatic data-processing.

---

[1] Principally, under Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3 (hereafter, AP I).

[2] Restated in Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law, Vol. 1: Rules* (CUP, 2005) (hereafter, CIHL). All Rules available at: <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul> accessed 10 June 2018.

[3] Yves Sandoz, Christophe Swinarski and Bruno Zimmermann (eds.), *Commentary on the Additional Protocols of 8 June, 1977, to the Geneva Convention of 12th August,1949* (Martinus Nijhoff, 1987) (hereafter, *AP I Commentary*).

[4] Program on Humanitarian Policy & Conflict Research at Harvard University (HPCR), *HPCR Manual on International Law Applicable to Air and Missile Warfare* (Harvard College, 2009) (hereafter, *AMW Manual*).

[5] HPCR, *Commentary on the HPCR Manual on International Law Applicable to Air and Missile Warfare* (v2.1, Harvard College, 2010) (hereafter, *AMW Manual Commentary*).

Subsequently, 7.3 examines precautionary measures, both passive (7.3.2) and active (7.3.3). Three themes will emerge here. First, active precautions will be taken in how LAWS are deployed and used, *and* LAWS deployments may in themselves represent a precautionary measure. Second, there may be an important interplay between active and passive precautions, in that announcing the deployment of LAWS may enhance the obligation of the opposing side to take certain passive precautionary measures. Third, the automatic processing capabilities of a LAWS are arguably more amenable to narrow precautionary measures, relative to proportionality, and this may enhance the ability of commanders to achieve the latter. This is complemented in 7.3.5, where it is argued that precautionary measures should be taken relatively earlier and in the broader context of military *operations*, not just attacks. Namely, the Joint Targeting process itself has pervasive precautionary value for LAWS deployments. Not least because it lends itself to a strong element of 'front-loading', and because it allows for a broader range of technically apt precautionary measures to be developed; a theme that will be explored in 7.3.6.

7.2 The Principle of Proportionality

The principle of proportionality protects civilians and civilian objects that are *not* directly targeted, but which may be *incidentally* affected in an otherwise legitimate strike. It is engaged after the commander satisfies the principle of distinction and takes all feasible precautions,[6] at which point the proportionality rule provides an "extra protective layer" of commensurability.[7] Specifically, it is prohibited to launch or continue with an attack that:

> …may be *expected* to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be *excessive* in relation to the concrete and direct military advantage *anticipated*.[8]

[6] *Prosecutor* v. *Galić* (ICTY Trial Judgment) IT-98-29-T (5 December 2003), ¶ 58; (Appeals Judgment) IT-98-29-A (30 November 2006), ¶ 190.

[7] Remarks by Janina Dill, 'Interpretive Complexity and the IHL Principle of Proportionality' (2014) 108 Proceedings of the Annual Meeting of the American Society of International Law 82, 83.

[8] Articles 51(5)(b), 57(2)(a)(iii) and (b), AP I (emphasis added). See also *AMW Manual*, Rules 14, 32(c) and 35(c); CIHL, Rules 14, 18 and 19.

In all, the concept is codified four times in AP I,[9] albeit without any reference to the term 'proportionality'. Instead, the focus is on ensuring that the expected collateral damage[10] (ECD) from an attack is not 'excessive' in relation to the military advantage anticipated (MAA).[11] In that regard, Schmitt notes that "[w]hile the rule is easily stated, there is no question that proportionality is among the most difficult of [LOAC] norms to apply".[12] Similarly, in its *Final Report*, the Office of the Prosecutor to the ICTY pointed out that:

> It is much easier to formulate the principle of proportionality in general terms than it is to apply it to a particular set of circumstances because the comparison is often between *unlike quantities and values*.[13]

This requirement to compare 'apples with oranges' introduces a strong element of subjectivity, which bedevils consensus in a given case or consistent application across cases. As Dill notes, "[l]oss of human life and military gain are never in harmony: reasonable people disagree on all but the most extreme cases of excessive civilian casualties".[14] Arguably, proportionality will present the single most difficult compliance problem for LAWS;[15] though, difficulty need not mean intractability.[16]

---

[9] (1) In Article 51(5)(b), as part of the prohibition on indiscriminate attack; (2) in Article 57(2)(a)(iii), as part of the precautionary considerations when launching an attack; (3) in Article 57(2)(b), for when an attack is in progress and may need to be cancelled or suspended; and (4) in Article 85(3)(b) on acts which, when done wilfully, are regarded as grave breaches of AP I.

[10] While neither 'collateral damage' nor 'collateral effects' are legal terms, they are used interchangeably here as shorthand for both incidental damage to civilian objects, and incidental death and injury to civilians.

[11] Note that CIHL, Rule 14, contains near-identical wording, but is entitled 'Proportionality in Attack'.

[12] Michael N. Schmitt, *Essays on Law and War at the Fault Lines* (TMC Asser Press, 2012), 190.

[13] Office of the Prosecutor, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia* (2000) 39 ILM 1257 (hereafter, *OTP Report*), ¶ 48 (emphasis added).

[14] Janina Dill, *Legitimate Targets? Social Construction, International Law and US Bombing* (CUP, 2015), 251. See also *OTP Report*, ibid., ¶ 50 (noting "it is unlikely that military commanders with different doctrinal backgrounds and differing degrees of combat experience or national military histories would always agree in close cases").

[15] Yoram Dinstein, 'Autonomous Weapons and International Humanitarian Law' in Wolff Heintschel von Heinegg, Robert Frau and Tassilo Singer (eds.), *Dehumanization of Warfare: Legal Implications of New Weapons Technologies* (Springer, 2018), ¶20 ("The principal problem that the [LOAC] will have to address upon baptism by fire of AI robots relates to the application and calculation of proportionality in attack"). See also Markus Wagner, 'Autonomy in the Battlespace: Independently Operating Weapon Systems and the Law of Armed Conflict' in Dan Saxon (ed.), *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff, 2013), 115 (noting that proportionality can be stated, but not fully defined in the abstract).

[16] See 7.2.3 on ensuring compliance of LAWS with the principle of proportionality.

Despite its misgivings, proportionality is a vivid example of the military necessity-humanity balance,[17] as it explicitly accepts the harsh reality of civilian harm, so long as this is 'justified' by the military advantage of attacking a lawful target; yet, it puts an upper limit on that harm, despite any perceived sense of military necessity by the attacker.[18] Accordingly, proportionality is one of several clear and categorical rejections of *Kriegsraison*.[19] That said, some pivotal terms codifying the principle are often ignored or misapplied, thus they are in need of clarification.

7.2.1 Clarifying Some Pivotal Terms

As emphasised in italics in the above reference to the AP I norm, the terms 'expected' and 'anticipated' make clear that the commander's judgment of the opposing variables (ECD and MAA) is made *before* an attack is launched. Thus, he is not dealing in certainties, but is making an assessment based on the information available at the relevant time.[20] Should this information be subsequently found to be flawed or incomplete, it is still the *ex ante* situation that matters for the legal assessment of proportionality.[21] In line with the *Rendulic Rule*,[22] the linchpin is foresight not hindsight;[23] thus, the popular trend towards effects-based condemnations is legally incorrect. This undoubtedly makes the principle more administrable to those on the ground, and to those seeking to deploy LAWS.

Crucially, the MAA must be "concrete and direct" in that it must be "substantial and relatively close" to the kinetic effects of an attack.[24] This does not mean the MAA should crystallise instantly,[25] but it does rule out putative advantages that are "hardly perceptible" and "which would only appear in the long term".[26] The *AMW Manual Commentary* takes all this to mean that the MAA should be "clearly identifiable and,

---

[17] *AP I Commentary*, ¶ 2206.

[18] Michael Newton and Larry May, *Proportionality in International Law* (OUP, 2014).

[19] *United States* v. *List (Wilhelm) et al. (The Hostage Case)* Case No. 7, 19 February 1948 (1950) 11 TWC 1230, 1256.

[20] William H. Boothby, *The Law of Targeting* (OUP, 2012), 94.

[21] Ibid., 95.

[22] *The Hostage Case*, 1297.

[23] Yoram Dinstein, *The Conduct of Hostilities Under the Law of International Armed Conflict* (3rd ed., CUP, 2016), 157; *AMW Manual Commentary*, Rule 14, ¶ 5.

[24] *AP I Commentary*, ¶ 2209.

[25] Ian Henderson, *The Contemporary Law of Targeting: Military Objectives, Proportionality and Precautions in Attack under Additional Protocol I* (Martinus Nijhoff, 2009), 200 (noting the conspicuous absence of 'immediate', or any similar wording in the AP I provisions).

[26] *AP I Commentary*, ¶ 2209.

in many cases, *quantifiable*".[27] This can also be seen to support (or at least to not undermine) LAWS deployments, as programmers and battle staffs will need to think through MAA values and consider their significance, before translating these into machine code.

A related matter concerns how closely confined to the current battle the MAA must be to qualify as 'direct'. The *AP I Commentary* suggests that proportionality is to be assessed on the tactical level,[28] and this has some qualified academic support.[29] The more accurate view, however, is that the MAA relates to the 'attack as a whole' and not from isolated or particular parts of it.[30] Not only does this have strong academic and expert support,[31] it is also reflected in military practice[32] and in the Rome Statute, which explicitly adds the adjective "overall" to its formulation of MAA.[33] Importantly, this broader view is also consistent with the idea that the MAA and proportionality are more appropriately assessed at the operational or strategic level. At these levels, senior commanders are able to consider the larger operational picture and come to a more comprehensive judgment.[34] Conversely, small-unit commanders and individual combatants often have a limited objective and are not briefed on its value within the overall scheme of attack, nor on its changing value as battlefield dynamics evolve.[35] Accordingly, such personnel should avoid unilaterally halting an attack, except in the case of extreme or unconscionable collateral damage. Once again, this may be seen to

---

[27] *AMW Manual Commentary*, Rule 14, ¶ 9 (emphasis added).

[28] *AP I Commentary*, ¶ 2207.

[29] Henderson (n 25), 202 ("…an attacker cannot argue 'this war is about national survival, therefore the allowable collateral damage is great'…However…attacks on strategic objectives [can] be assessed in [their strategic] context").

[30] UK, Reservations and Declarations Made Upon Ratification of AP I (28 January 1998), Statement (i).

[31] Boothby (n 20), 95; Schmitt (n 12), 192; Dinstein (n 23), 161; *AMW Manual Commentary*, Rule 14, ¶ 11.

[32] See 5.3.1.2 on Target System Analysis, which recognises the substitutability of some targets (from the enemy's point of view), hence the need to attack them together, for the military advantage to materialise. See also *AMW Manual Commentary*, Rule 14, ¶¶ 12-13 (referring to multiple bridges across the same river and attacks as ruses).

[33] Article 8(2)(b)(iv), Rome Statute of the International Criminal Court (adopted 17 July 1998, entered into force 1 July 2002) 2187 UNTS 3; UN Doc. A/CONF.183.9. As Boothby (n 20), 97, points out, despite this provision fulfilling its own purpose under international criminal law (ICL), it is still worth bearing in mind here as it represents the most recent internationally adopted text on proportionality.

[34] Dinstein (n 23), 108 and 161; W. Hays Parks, 'Air War and the Law of War' (1990) 32 Air Force Law Review 1, 172 and 175-76.

[35] Hays Parks, ibid., 176 (citing the extreme example of the D-Day landings, where only a top General such as Dwight D. Eisenhower was in a position to measure the MAA of each component of the bombing attack, thus it would be inappropriate for an individual pilot to halt his own attack in response to a few extra civilians appearing).

support LAWS deployments, as tactical autonomous units will be unlikely to undertake their own proportionality assessment for every strike, but may instead proceed with pre-programmed thresholds and/or real-time human input,[36] in line with the 'individual attack' limitation.[37]

As alluded to above, the biggest problem with the proportionality principle is its subjective application and the idea of 'weighing' dissimilar values.[38] On the one hand, collateral damage can now be estimated with a high level of rigour and objectivity, by using the Collateral Damage Estimation Methodology (CDEM).[39] On the other hand, the MAA of an attack is not only subjective,[40] but is highly contextual and, therefore, constantly shifting as the battlefield evolves.[41] When 'weighing' the two variables, the fulcrum word that should guide any assessment is 'excessive' and not 'disproportionate'. The latter suggests that a fine balancing of the two is required and that the rule is violated with a slight tipping of the scales.[42] Yet, despite some prominent academic[43] and judicial[44] support for this view, notions of 'disproportion' do not actually appear anywhere in the text of AP I. Instead, the proportionality rule is framed as an 'indiscriminate attack', and its negotiating history shows that

---

[36] See 7.2.3.

[37] See 4.3 and 5.5.2.

[38] Dill (n 14), 84 (noting that "human life and military gain cannot be expressed or determined in terms of each other" and that "a gain in one [usually] implies a loss in the other", such that finding a legally correct 'balance' between these two often competing variable is "prima facie subjective").

[39] This US-based methodology takes certain inputs, such as the precision of a weapon; its blast radius; other known materials, objects and explosives within that radius; attack tactics; and the probability of civilian presence near targets. These are combined to derive a 'collateral effects radius' and 'potential collateral damage' output figure, comprising both civilians and civilian objects. See Jeff Thurnher and Tim Kelly, 'Panel Discussion: Collateral Damage Estimation', *US Naval War College* (23 October 2012) <https://www.youtube.com/watch?v=AvdXJV-N56A> accessed 6 September 2018.

[40] Dinstein (n 23), 162 (arguing that MAA is usually the progeny of military planning and may not be transparent to the external observer; thus, it cannot be assessed independently from the attacker's subjective state of mind).

[41] Newton and May (n 18); Wagner (n 15), 112.

[42] WJ. Fenrick, 'Targeting and Proportionality During the NATO Bombing Campaign Against Yugoslavia' (2001) 12 European Journal of International Law 489, 501 ("…resolution of the proportionality equation requires a determination of the relative worth of military advantage gained by one side and the civilian casualties or damage to civilian objectives incurred in areas in the hands of the other side").

[43] For example, ibid.; WJ. Fenrick, 'The Rule of Proportionality and Protocol I in Conventional Warfare' (1982) 98 Military Law Review 91, 106 (suggesting there is an equivalence between 'excessive' and 'disproportionate').

[44] *Public Committee against Torture in Israel* et al. v. *Government of Israel* et al. (2005) HCJ 769/02 (*Targeted Killings Case*), ¶ 45 (referring to a "balancing between conflicting values and interests", so that the benefit arising from an attack on a military objective is "proportionate to the damage caused to innocent civilians").

'disproportionate' was deliberately replaced with 'excessive'.[45] This was to mitigate the subjective element and the 'apples and oranges' problem of needing to devise an 'exchange value' between incommensurate variables.[46] Accordingly, a *significant* and unreasonable outweighing of MAA by ECD is needed before the rule is violated.[47] Complementing this approach, the commander enjoys a "fairly broad margin of judgment",[48] the combined effect of which is to make the rule more administrable and to decrease the likelihood of inadvertently violating it. Notably, the Rome Statute affords an even greater margin of appreciation, with the adverb "clearly"[49] implying that the disparity between ECD and MAA must be "gross and obvious" before an offence has been committed.[50]

Significantly, this leads to the argument that proportionality is not necessarily subjective, but is an *objective* standard[51] that is *applied* in a subjective manner because fallible human beings are conducting the assessments.[52] This is arguably in line with the ICTY Trial Chamber's judgment in *Galić*, which articulated the 'reasonable military commander' standard as:

> …whether a *reasonably well-informed person* in the *circumstances of the actual perpetrator*, making *reasonable use* of the information available to him or her, could have expected excessive civilian casualties to result from the attack.[53]

---

[45] For a summary of the diplomatic woes over the wording, see Hays Parks (n 34), 171-72; Samuel Estreicher, 'Privileging Asymmetric Warfare (Part II)?: The "Proportionality" Principle under International Humanitarian Law' (2011) 12 Chicago Journal of International Law 143, 151-53.

[46] Ibid. See also *Prosecutor* v. *Galić* (Trial), where the ICTY Trial Chamber did not distinguish between knowingly causing excessive collateral damage and a careless indiscriminate attack (¶ 387), as both abandon the legal requirement to spare civilians as much as possible (¶ 58).

[47] Schmitt (n 12), 190; Dinstein (n 23), 158-59; Boothby (n 20), 97; *AMW Manual Commentary*, Rule 14, ¶ 7. See also the *Targeted Killings Case*, ¶ 58 (discussing the "zone of proportionality").

[48] *AP I Commentary*, ¶ 2210.

[49] That is, collateral damage must be "clearly excessive in relation to the concrete and direct overall [MAA]" under Article 8(2)(b)(iv), Rome Statute.

[50] Boothby (n 20), 97.

[51] Enzo Cannizzaro, 'Proportionality in the Law of Armed Conflict' in Andrew Clapham and Paola Gaeta (eds.), *The Oxford Handbook of International Law in Armed Conflict* (OUP, 2014), 337 ("By nature, proportionality is an objective standard. There would be no use in requiring a State to consider humanitarian interests in the pursuit of its military objectives if the balance of interests could be struck according to the State's subjective perception"). See also *AMW Manual Commentary*, Rule 14, ¶ 6.

[52] Newton and May (n 18); Marco Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified' (2014) 90 International Law Studies 308, 331-33.

[53] *Prosecutor* v. *Galić* (Trial), ¶ 58 (emphasis added).

A ‘reasonably well-informed person’ who makes ‘reasonable use’ of the available information is clearly an objective standard; qualifying this with the ‘circumstances of the actual perpetrator’ allows for a realistic subjective element.[54] As further clarification, and partly in line with *Galić,* Wright argues for an administrable two-part ‘subjective-objective’ test.[55] For part one, the commander’s evaluation of both the ECD and MAA is the result of a *subjective assessment*, done with an *objective* degree of diligence: namely, with “common sense and good faith”, and in light of the available information.[56] For part two, these opposing values are compared for an *objective determination* of whether the ECD was ‘proportionate’, ‘excessive’ or ‘clearly excessive’.[57] On its face, this approach is potentially amenable to the automatic processing capabilities of a LAWS, which can reliably routinise the comparison of data values; subject, however, to human input on the more deliberative aspects of the initial valuations.

Taking the ‘objective standard’ argument one step further, Estreicher argues that the “proper test [for excessiveness] is one of ‘necessity’”.[58] Namely, so long as the attack does in fact offer a concrete and direct MAA,[59] the commander simply needs to select the “least deleterious (in terms of civilian loss) means of achieving that objective”, without needing to inquire into “complex, metaphysical exchange rates” between dissimilar values.[60] This interpretation is clearly more aligned with military operational and tactical logic,[61] and it is undoubtedly the one that is most amenable to the automatic processing capabilities of a LAWS.[62] However, it is also legally incorrect as it focuses exclusively on distinction and precautions in attack.[63] Proportionality has an independent (and somewhat subjective) existence beyond the

---

[54] Henderson (n 25), 222 (noting that after the subjective evaluation of MAA, “the *conclusions* to be reached on whether collateral damage is *expected* and whether it is *proportional*” are objective standards) (original emphasis).

[55] Jason D. Wright, ‘Excessive Ambiguity: Assessing and Refining the Proportionality Standard’ (2012) 94 International Review of the Red Cross 819, 851-52.

[56] *AP I Commentary*, ¶ 2208.

[57] Wright (n 55), 851-52. Note that ‘clearly excessive’ is specific to ICL proceedings.

[58] Estreicher (n 45), 156.

[59] And holding that MAA constant between the different means, methods and objects of attack.

[60] Estreicher (n 45), 156-57.

[61] In that it offers a more objective and administrable standard for commanders and their battle staffs to apply.

[62] In that it merely requires the machine the recognise a military objective, before selecting the munition, and the timing and direction of attack that minimises the ECD value from its integrated CDEM.

[63] In particular, Articles 52(2) and 57(2)(a)(ii), AP I.

precautionary measures; the latter can greatly mitigate the difficulty for commanders when assessing the former,[64] but the two are not necessarily interchangeable.[65]

Finally, the *AP I Commentary* seems to assume a 'finite relationship' between ECD and MAA, as it categorically rules out "extensive" collateral damage.[66] Ostensibly, this would require a finer level of judgment, thereby adding to the difficulty of the 'balancing exercise'. Yet the weight of academic and expert opinion is equally categorical: that the fulcrum word is 'excessive'; that this continues to govern the relationship between ECD and MAA, even at very high levels of both variables; and, therefore, that neither 'extensive' nor even 'severe' collateral damage is automatically unlawful.[67] In short, 'context is king' and even horrendous civilian casualties may be lawful if the MAA is significant enough to 'justify' these.

7.2.2 Problematic Compliance with Proportionality in LAWS Deployments

First, note that proportionality is about limiting the harm to *civilians* and has nothing to do with the effects of armed conflict on combatants.[68] Thus, similar to the principle of distinction, remote battlefields and demilitarised zones which contain no civilians should pose no proportionality issue. Conversely, where civilians *are* present in or near the contact zone, proportionality will arguably pose similar and even *greater* challenges for LAWS than distinction.[69]

To elaborate, the ability to distinguish between civilians and combatants (and between civilian and military objects) is a prerequisite for LAWS to be able to undertake proportionality assessments;[70] namely, by correctly sorting persons and objects into either the ECD or MAA category. Yet it does not end there: while the CDEM can

---

[64] *AMW Manual Commentary*, Rule 14, ¶ 5. See also 7.3.5.

[65] As Dill (n 14) notes at 84-85, there are three criteria to operationalise proportionality: adequacy (distinction); necessity (minimising collateral damage); and commensurability (the final protective layer). Thus, Estreicher clearly stops short of giving full legal effect to the proportionality principle.

[66] *AP I Commentary*, ¶ 1980.

[67] For example, Dinstein (n 23), 156-57; Newton and May (n 18), 164; Wagner (n 15), 118; Schmitt (n 12), 191; *AMW Manual Commentary*, Rule 14, ¶ 8.

[68] Dinstein, ibid., 154-55.

[69] Wagner (n 15), 115.

[70] Wagner, ibid., 119.

potentially be automated[71] and programmed to recognise 'any non-positively identified person or object' as being of a civilian nature,[72] the overwhelming weight of academic, expert and NGO opinion is that LAWS will struggle to assess the MAA.[73] This is because the MAA of an attack is both context-specific and holistic,[74] and it constantly evolves, depending on the commander's plans and the development of military operations on both sides.[75] Certainly, attacking a command and control post at the start of an armed conflict will yield greater military advantage than towards the end, when enemy forces are in disarray and nearing defeat.[76] Exactly *how much* higher or lower – or indeed whether the MAA remains 'concrete and direct' – is a contextual question, which demands consideration of a broad range of *qualitative* factors.[77] In turn, this would:

> …necessitate understanding of military strategy, operational issues, and tactics. The LAWS would furthermore have to be able to comprehend continual changes in goals and objectives, internal changes to their relative importance, and the anticipated military utility of achieving them.[78]

---

[71] Michael N. Schmitt and Jeffrey S. Thurnher, 'Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict' (2013) 4 Harvard National Security Journal 231, 255 ("[CDEM] analysis is performed using objective data and scientific algorithms", therefore it is clearly amenable to automatic processing).

[72] In line with the negative definitions of civilian and civilian object in Articles 50(1) and 52(1), AP I, respectively.

[73] For example, Schmitt and Thurnher (n 71), 255; Sassóli (n 52), 331-32; William H. Boothby, *Conflict Law: The Influence of New Weapons Technology, Human Rights and Emerging Actors* (TMC Asser Press, 2014), 110; Nathalie Weizmann, 'Autonomous Weapon Systems under International Law', *Academy Briefing No. 8* (Geneva Academy of International Humanitarian Law and Human Rights, November 2014), 15; Robin Geiss, *The International Law Dimension of Autonomous Weapons Systems* (Friedrich Ebert Stiftung Study, October 2015), 15; Human Rights Watch, *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch, 2012), 32-34.

[74] Pablo Kalmanovitz, 'Judgment, Liability and the Risks of Riskless Warfare' in Nehal Bhuta et al. (eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (CUP, 2016), 150-51 (pointing out that CDEM is primarily a policy tool for determining the level of command required to authorise an attack. The greater the collateral risk, the higher the authority needed for clearance, precisely *because* the assessment of MAA is for a human commander to make. Namely, the more potential harm there is for protected persons and objects, the more judgment and experience the commander will need when deciding whether such harm is 'excessive' in relation to the concrete and direct MAA).

[75] Sassóli (n 52), 331-32.

[76] Schmitt and Thurnher (n 71), 255.

[77] Wagner (n 15), 120; Markus Wagner, 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems' (2014) 47 Vanderbilt Journal of Transnational Law 1371, 1397.

[78] Kjølv Egeland 'Lethal Autonomous Weapon Systems under International Humanitarian Law' (2016) 85 Nordic Journal of international Law 89, 104; though, see 7.2.3.

Accordingly, the mental operation required to deliberate the MAA calls for "complex, value-based *case-by-case* decision-making in which circumstances have to be weighed in their totality".[79] This all points to a clear geographical and temporal dimension, in that the MAA may be accurately known on deployment, but the longer a LAWS loiters and the further out it travels, the more the battlespace – hence, the MAA of a given attack – is liable to change. According to Sassóli, this is "the most serious IHL argument" against even the theoretical possibility of prolonged autonomy, and it can only be resolved if the system is "constantly updated about military operations and plans".[80]

On the other hand, there are two counter-arguments. First, longer loitering times and superior sensory capabilities may afford an opportunity to wait for a more 'proportionate' situation to arise before a LAWS executes a specific attack.[81] Second, recall the *Rendulic Rule* and the focus of the proportionality standard on *ex ante* judgments. In this connexion, the legally relevant proportionality assessment may be the one undertaken by the commander on deployment.[82] This is similar to the case of launching a tactical cruise missile, where any unforeseen changes in ECD and MAA are largely irrelevant to the initial decision to deploy,[83] and it would most likely apply to a LAWS deployed on a targeted strike.[84]

A large body of academic, expert and NGO opinion also considers that the 'weighing' exercise will pose even greater problems for LAWS. To be sure, assessing proportionality will rarely be a case of counting civilian bodies *versus* combatant bodies;[85] more often, the challenge is to determine the number of acceptable civilian casualties in exchange for the destruction of a tank or a bridge.[86] As mentioned above, even if correct values are known, the incommensurability between ECD and MAA

[79] Geiss (n 73), 15 (emphasis added).

[80] Sassóli (n 52), 332. In this regard, see the alternative options in 7.2.3.4 *versus* 7.3.6.4-7.3.6.6.

[81] Wagner (n 77), 1413-14. For example, unlike a 'fire-and-forget' munition, a LAWS can loiter until passing civilians have moved out of the collateral effects radius; or it can follow a military objective until the latter moves into a deserted area, where a kinetic attack would inflict lower, or no collateral damage.

[82] William H. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' in Dan Saxon (ed.), (n 15), 58.

[83] Ibid., 59.

[84] See 7.2.3.

[85] Schmitt and Thurnher (n 71), 254; *AMW Manual Commentary*, Rule 14, ¶ 7.

[86] Dill (n 14), 84; Wagner (n 15) 120.

negates any "reasonably exact proportionality equation between them";[87] hence, no universally accepted or encodable definition of 'excessive' exists.[88] Instead, the proportionality assessor has to consider the full variety of available data *and* assign relative weights to each aspect in the circumstances ruling at the time,[89] and on three distinct levels.[90]

At this point, it is worth recalling the distinction between rules and standards from 3.2.2.2.1. Arguably, proportionality is *the* canonical example of a standard that calls for the most deliberative and metacognitive thinking in its application.[91] This is less to do with any apparent subjectivity, and more a result of the need for case-by-case determinations, taking into account numerous variables and infinite combinations thereof; and resolving unique value conflicts that will not have been foreseen, much less coded in advance. As Cannizzaro argues, this means:

> [P]roportionality can be seen as a *law-making process* that continuously adapts the content of the rule to changing social needs, circumventing the complexities of the rules of change in the international legal order.[92]

To be sure, machines may *apply* well-defined rules that are amenable to automatic processing,[93] but they cannot *make* legal rules.[94] The latter requires controlled processing to perceive competing interests, assess their respective weights in the light of specific circumstances, and to compare one with the other, before formalising a new rule.[95]

---

[87] Fenrick (n 43), 102.

[88] Dill (n 14), 84; Schmitt and Thurnher (n 71), 254; Boothby (n 73), 110; Weizmann (n 73), 15; Human Rights Watch (n 73), 36.

[89] Markus Wagner, 'Taking Humans Out of the Loop: Implications for International Humanitarian Law (2011) 21 Journal of Law, Information & Science 155, 163.

[90] Wagner (n 77), 1399 (namely, target selection; choice of munition; and the means of engagement, such as the direction of attack. This broadly reflects the Joint Targeting process detailed in 5.3).

[91] Cannizzaro (n 51), 332-33 (contrasting the 'classical model' of law-making, where "the normative content is determined directly by the rule", and the 'proportionality model', where "the normative content emerges…from a secondary process of law-making based on the assessment of proportionality").

[92] Ibid., 333 (emphasis added).

[93] Recall from 3.2.2.2.1 that this includes playing chess and Go, and even applying road-traffic law, much of which consists of well-defined rules that impose strict liability.

[94] Human Rights Watch (n 73), 32-33 (noting that it would be impractical to program a LAWS with all the situations and combinations of value conflicts in advance; or, to expect that it will fully understand its environment and perceive every relevant cue required to apply the proportionality standard).

[95] Ibid., 33 ("Those who interpret [LOAC] in complicated and shifting scenarios consistently invoke human judgment, rather than the automatic decision-making characteristics of a computer").

Yet, given the above discussion on 'disproportionate' *versus* 'excessive', this problem should not be exaggerated on account of autonomy. NGO claims that a LAWS would have to engage in a "*delicate balancing* of the two factors [of ECD and MAA]"[96] and would be "unlikely to be able to qualitatively *balance* them"[97] are arguably raising the proportionality challenge to a difficulty level that even humans cannot routinely meet. Instead, the correct standard is one of 'excessiveness', whereby the objectively reasonable commander would see the ECD as *significantly* higher than the notional MAA. This makes it relatively more likely that a LAWS would act 'proportionately' on the battlefield, given appropriate real-time human input and/or advance precautionary measures. That said, the importance of these latter two (alternative) conditions should not be underestimated: given the antecedent risk of machines wrongly evaluating the MAA, the problem of "garbage in = garbage out" will always bedevil the *fully* autonomous proportionality assessment.[98] Again, we arrive at the conclusion that the assessment of MAA and proportionality calls for strong deliberative reasoning and a distinctly human judgment, which must be rendered on a case-by-case basis. As will be seen below, this can occur in a LAWS context, but only with appropriate human input before and/or during deployment.

To sum up, despite some administrable features, the proportionality principle as a whole remains problematic for LAWS deployments for three reasons.

- The *shifting* and *contextual* nature of the MAA, many aspects of which are likely to be abstract and not amenable to machine perception.
- The need to compare *contradictory* and *dissimilar* values with *no common metric*; namely, ECD *versus* MAA. This creates an 'apples and oranges' problem that may be stated in general terms, but cannot be precisely defined in the abstract, let alone in software code.
- The *inherently indeterminate* nature of 'excessive', which also makes the application of proportionality highly contextual, though subject to a standard of reasonableness within a broad margin of judgment.

---

[96] Human Rights Watch, *Advancing the Debate on Killer Robots: 12 Key Arguments for a Preemptive Ban on Fully Autonomous Weapons* (Human Rights Watch, May 2014).

[97] Ibid., 6 (emphasis added).

[98] HR. Taylor, *Data Acquisition for Sensor Systems* (Springer, 1997), 3.

The importance of the above barriers should not be underestimated. As noted in 2.2.3.2, robots need tasks and desired/prohibited actions to be both *precise* (specified in sufficient programmable detail) and *tangible* (with quantifiable expected outcomes), if they are to operate autonomously, reliably and lawfully.[99] Raw "common sense and good faith"[100] will very likely elude a LAWS, and will lead to unlawful actions on the battlefield; unless there is timely and appropriate human input, to translate these into machine-perceptible data.

7.2.3 Ensuring the Compliance of LAWS Operations with the Principle of Proportionality

7.2.3.1 Restricted Deployments

As noted above, proportionality – like distinction – will not pose any legal challenge where there is little or no civilian presence. Accordingly, one solution that is implicit in much of Wagner's writings is to restrict LAWS deployments to remote battlefields, to attack military objectives by nature.[101] This option is also supported by Thurnher, who discusses high-intensity conflict in remote deserts or undersea, or deployments in demilitarised zones.[102] Boothby echoes this with an additional restriction: to deploy LAWS in a "remote, unpopulated or sparsely populated area [to search] for *specific* military objectives whose destruction would be militarily *most valuable*".[103] Namely, to use LAWS in these benign environments specifically for targeted strikes (TS) against high-value targets (HVT), for the greatest possible control and the lowest risk of 'disproportionate' attack.

Crootof adds a similar but slightly wider possibility, suitable for tactical-level combat (TLC). That is, where the commander has determined that *all foreseeable* engagements within a tight set of spatio-temporal restrictions and specific target parameters[104] comply with the proportionality principle, and he duly authorises the LAWS to select

[99] Sassóli (n 52), 331

[100] *AP I Commentary*, ¶ 2208.

[101] Wagner (n 89), 163; (n 15), 122.

[102] Jeffrey S. Thurnher, 'No One at the Controls: Legal Implications of Fully Autonomous Targeting' (2012) 67 Joint Force Quarterly 77, 80.

[103] Boothby (n 82), 57 (emphasis added).

[104] See also 7.3.6.4 and 7.3.6.5.

and engage targets within those constraints.[105] Indeed, the *Harpy* is currently employed on this basis, with commanders doing *ex ante* proportionality assessments when they are able to take into account anything that *may* reasonably occur during deployment.[106]

With the return to Great Power Competition,[107] the 'restricted deployments' option is arguably not as limiting as it might seem,[108] and it may well become more prominent over time. However, there is also likely to be a range of other situations in which States will want to deploy LAWS. For these situations, militaries will need to combine automatic and controlled processing in ways to effectuate a meaningful human control (MHC) and to maximise the likelihood of lawful outcomes on the battlefield. Some of the following deployment options will incorporate Newton and May's 'rules of thumb', which are intended to guide military planners "in a class of common cases in which proportionality considerations arise".[109]

7.2.3.2 Thurnher: 'Workarounds'

Thurnher has presented a number of 'workarounds' that may address the proportionality challenge in the face of civilian presence. First, the author credits future LAWS with integrated CDEM as being able to determine whether collateral damage exceeds a predetermined limit, but also acknowledges that the proportionality test calls for a greater sense of what is 'excessive'.[110] The workarounds then aim to exploit the data-processing advantage of CDEM, while shifting the onus onto the commander for controlled processing. These include the following.

- Deploying LAWS in TSs where high ECD is acceptable, for example, against HVTs.[111] This option would seem to invoke Newton and May's *Civilian Precautionary Principle*, which asserts, "[w]henever civilian lives are greatly risked…very clear weighty military objectives have to be discerned for the tactic to be *prima facie* proportionate".[112] Presumably, such situations would

[105] Rebecca Crootof, 'The Killer Robots are Here: Legal and Policy Implications' (2015) 36 Cardozo Law Review 1837, 1878.

[106] Recall from 2.4 that the *Harpy* only engages radar-emitting objects, within very tight spatio-temporal limits.

[107] See 1.2.2.

[108] Cf. Wagner (n 89), 163; (n 15), 122 (presenting this option as a necessary drawback, rather than a solution).

[109] Newton and May (n 18), 12.

[110] Thurnher (n 102), 82.

[111] Ibid., 81.

[112] Newton and May (n 18), 286.

also require the commander to have an idea of the likely concentration of civilians, in order to assess the risk of ECD rising beyond the MAA.

- Designating ECD thresholds per TS mission or per target engagement, where the commander knows the MAA in advance and is reasonably confident that this will not change significantly.[113] Notably, some HVTs (like enemy headquarters) are likely to remain fairly static in their MAA, certainly for the duration of a LAWS deployment.[114] The LAWS can then loiter until its CDEM reveals that ECD is below the threshold. Where the ECD exceeds the threshold, a LAWS can be programmed to hold fire; launch non-lethal munitions, to get civilians and vehicles to flee the collateral effects radius; or it can seek human approval for the engagement.[115]
- In the case of TLC, changes in MAA can even be pre-programmed in accordance with known or anticipated changes on the battlefield, such as how many other targets have been destroyed or neutralised;[116] or whether tanks appear individually or in concentrations.[117] Importantly, such *ex ante* determinations need not be perfect, so long as they are reasonable in the circumstances. This is because commanders enjoy a "fairly broad margin of judgment"[118] in their proportionality assessment, which must reflect "common sense and good faith".[119]
- In support of the above, some additional control measures may be prudently employed. For example, tight spatio-temporal constraints will reduce the risk of MAA changing beyond the commander's reasonable expectation.[120] In addition, as MAA is contextual, conservative ECD thresholds can be set with a dial-in capacity for commanders to alter these further, as and when required.[121]

---

[113] Thurnher (n 102), 81.

[114] Lieutenant Colonel Christopher M. Ford, 'Autonomous Weapons and International Law' (2017) 69 South Carolina Law Review 413, 445 ("Even on today's modern, fast-moving battlefield, the military advantage of some targets remains fairly static").

[115] Thurnher (n 102), 81. See also 7.2.3.3.

[116] See also 7.2.3.4.

[117] Schmitt and Thurnher (n 71), 256-57.

[118] *AP I Commentary*, ¶ 2210.

[119] Ibid., ¶ 2208.

[120] Jeffrey S. Thurnher, 'Means and Methods of the Future: Autonomous Systems' in Paul AL. Ducheine, Michael N. Schmitt and Frans PB. Osinga (eds.), *Targeting: The Challenges of Modern Warfare* (TMC Asser Press, 2016), 189. See also 7.3.6.4.

[121] Schmitt and Thurnher (n 71), 256. See also 7.2.3.4.

7.2.3.3 Boothby: The Precautionary Value of Process

Boothby considers a more data-driven approach, which strongly integrates precautions in attack.[122] Of particular importance here is 'pattern of activity'[123] data to establish when and where large civilian movements typically take place, and to derive a reasonable prediction of the collateral risk of attacking a particular target (or target group) at a particular time.[124] This will help to determine the exact time of attack, to minimise both ECD and the likelihood that it may be excessive. Yet, unexpected events may occur, such as a column of refugees crossing a previously deserted area. In such a case, there are two possible solutions.[125]

- A sophisticated LAWS may detect the presence of more civilians than expected, wait for them to pass and to clear the area before initiating an attack; or its mission control software may suspend the search and contact commanders for a proportionality reassessment.[126]
- Alternatively, a less sophisticated LAWS may simply be fed with an expected image of the contact zone before deployment. Its mission control software can then be programmed to refrain from any attack if the image observed on the battlefield differs in any material respect from the programmed image.[127]

7.2.3.4 Van den Boogaard: Proportionality and the Levels of Warfare

Arguably the most sophisticated solution, which also combines automatic and controlled processing, comes from van den Boogaard, who focuses on the level of command at which proportionality assessments should take place.[128] Recall from 7.2.1 that for practical reasons, proportionality has traditionally been assessed at the strategic or operational level, while the *AP I Commentary* requires it to be assessed at the tactical level. Similarly, LAWS are assumed to only be able to determine the ECD and MAA of their own attack (tactical), while State Practice is to consider these from the attack

---

[122] Boothby (n 82), 56-58. See also 7.3.

[123] 'Pattern of activity' describes the aggregate behaviours and activities of all entities (persons, vehicles, etc.) flowing through a given geographic location, as well as their timings. See Patrick Biltgen and Stephen Ryan, *Activity-Based Intelligence: Principles and Applications* (Artech House, 2016), 116 and 119-20.

[124] Boothby (n 82), 58.

[125] Both of which potentially comply with the 'cancel or suspend' obligation in Article 57(2)(b), AP I. See 7.3.2.3.

[126] Boothby (n 73), 118.

[127] William H. Boothby, 'Autonomous Attack – Opportunity or Spectre?' in Terry D. Gill (ed.), *Yearbook of International Humanitarian Law 2013*, Vol. 16 (TMC Asser Press, 2015), 83.

[128] Jeroen van den Boogaard, 'Proportionality and Autonomous Weapons Systems' (2015) 6 Journal of International Humanitarian Legal Studies 247, 275-77. On the levels of command, see 5.2.3.

as a whole (operational and strategic).[129] Accordingly, both ECD and MAA should ideally be estimated on all three levels.[130]

With this in mind, van den Boogaard envisages a TLC scenario with strong communication links and elaborate data flows: both *horizontally*, between tactical autonomous units; and *vertically*, between those same units and the operational and strategic headquarters.[131] In such a case, it is possible to continuously update the situation on each level, thereby deriving time-sensitive tactical, operational and strategic proportionality assessments. In particular, multilateral data flows and feedback enable each level of command to update the assessments of ECD, MAA and, therefore, 'proportionality' – all in *real time*. For example, if an attack by a tactical autonomous unit derives a military advantage and gets closer to achieving the goals of the larger operation, this will decrease the assigned value of the MAA that other parts of that same operation are intended to achieve, thereby reducing the likelihood that a 'disproportionate' attack may be planned at the operational level.[132] Likewise, if changes in strategic priorities occur at the political level, these can be manually inputted at strategic headquarters, thereby updating the MAA of targets assigned to each tactical unit; importantly, the human-inputted MAA must be 'converted' into an ECD threshold, for automatic processing. Accordingly, human commanders remain in control at the broader *operational* and *strategic* levels, due to the need for highly qualitative assessments involving strategic and political factors at those levels.[133] Concurrently, rapid and data-heavy proportionality calculations are autonomously done at the *tactical* level, which can sometimes move too fast for human operators to keep pace.[134] This enables (tactical) 'narrow loop' autonomy to proceed alongside human judgment in the 'wider loop' of (strategic and operational) control.

[129] Ibid., 275-76.

[130] Ibid., 276.

[131] Ibid., 276.

[132] Ibid.

[133] Ibid., 277.

[134] To reiterate: autonomous proportionality calculations here are derived from: *human-inputted* MAA data; conversion of that data to an ECD threshold; and actual ECD assessments done by each tactical LAWS unit.

This approach deftly addresses Sassóli's objection to prolonged autonomy in 7.2.2, above,[135] using a combination of a dial-in capacity and Newton and May's *Common Denominator Principle*.[136] It also has intuitive appeal and implicit support from other commentators,[137] yet it poses at least one major limitation: the system would require reliable and continuous communication links. This may be possible in an asymmetric (low-intensity) conflict, where the LAWS-deploying side will have a major technological advantage, but it cannot always be guaranteed. Indeed, as noted in 2.5.4.4, symmetric (high-intensity) battlefields – which will likely characterise conflicts resulting from Great Power Competition – often become GPS-denied or jammed environments. In such instances, communication links are either unavailable, or they must be deliberately switched off to avoid spoofing, and this will clearly undermine any remote dial-in capacity. That said, three recent developments in US defence circles are worth noting.

- Advances in anti-jam receiver technology are now providing over 10,000 times improved jamming resistance over previous models.[138]
- Next-generation wideband high frequency (WBHF) communication links have recently been refined, tested and demonstrated to transfer data files of up to one megabyte; both rapidly and reliably over a 5,000-mile distance.[139] The project is specifically intended to enable Air Force communications in GPS-denied and jammed environments, using robust terrestrial-based 'Beyond-Line-of-Sight' systems.[140]

---

[135] Namely, that the only way to address the shifting and contextual nature of the MAA is for systems to be "constantly updated about military plans and operations": Sassóli (n 52), 332.

[136] This seeks to address the incommensurability problem with the maxim: "[t]ry to find a common metric to translate both types of value". See Newton and May (n 18), 285 (suggesting the example of military objectives expressed in terms of lives saved, compared with civilian casualties).

[137] For example, Kalmanovitz (n 74), 151 (discussing the linkages between tactical and strategic military advantage: how LAWS may be able to estimate the former; and why the latter implicates political goals, thereby necessitating some human involvement in proportionality calculations).

[138] 'Rockwell Collins Delivers Latest Digital GPS Receiver Technology to US Air Force Special Operations Command', *Rockwell Collins News* (24 August 2017) <https://www.rockwellcollins.com/Data/News/2017-Cal-Yr/GS/20170824-DIGAR-delivery.aspx> accessed 8 September 2018.

[139] 'Rockwell Collins Successfully Demonstrates 5,000 mile Next-Generation Wideband High Frequency Communications Link', *Rockwell Collins News* (18 September 2017) <https://www.rockwellcollins.com/Data/News/2017-Cal-Yr/GS/20170918-PACAF-HF-Comms.aspx> accessed 8 September 2018.

[140] 'Wideband High Frequency Communications Provide Net-Centric, High-Speed Beyond Line of Sight Communications in Anti-Access Area-Denial (A2/AD) Battlefield Environments', *International Defence, Security & Technology* (11 August 2017) <http://idstch.com/home5/international-defence-security-and-technology/technology/ict/wideband-high-frequency-communications-provide-net-centric-high-speed-beyond-line-sight-communications-anti-accessarea-denial-a2ad-battlefield-environments/> accessed 8 September 2018.

- The US Army recently began testing and refining 'pseudolites' with human soldiers on the ground.[141] Pseudolites are a localised communication and PNT[142] infrastructure, consisting of satellite-like transmitters and powerful anti-jam antennas. The system functions like GPS, and is intended to augment or replace GPS signals when these are weak or inaccessible.[143] Transmission hardware is terrestrially located – either airborne or ground-based – and while these can be moved around, they are restricted to a defined geographical area at any one time. Thus, transmitters are situated closer to their intended receivers, enabling far stronger signals than is possible with GPS, hence there is significantly less scope for interference or deliberate jamming.

In a more distant future, LAWS may eventually become part of a 'giant, armed nervous system' in which "[e]very weapon, vehicle, and device [is] connected, sharing data, constantly aware of the presence and state of every other node in a truly global network".[144] At present, these plans are confined to the US military, and they remain both classified and at the conceptual stage.[145] However, tangible steps have been taken towards realising them with the Pentagon's recent announcement of the Joint Enterprise Defense Infrastructure (JEDI).[146] Together with the aforementioned projects, this may be expected to provide a robust and secure dial-in capability. Thus, there may be less reason in future to doubt the reliability of LAWS communication links, so long as deploying forces remain one step ahead of enemy counter-measures, and ensure their own infrastructure remains fit for purpose.

---

[141] Kathryn Bailey, 'Pseudolites Preserve Position Information During GPS-Denied Conditions', *US Army Press Release* (2 June 2016) <https://www.army.mil/article/169033/pseudolites_preserve_position_information_during_gps_denied_conditions> accessed 8 September 2018.

[142] That is, 'Position, Navigation and Timing', which is the essence of all GPS capabilities.

[143] Michael Jones, 'Army Pseudolites: What, Why and How?' *GPS World* (9 August 2017) <http://gpsworld.com/army-pseudolites-what-why-and-how/> accessed 8 September 2018.

[144] Patrick Tucker, 'The Future the US Military is Constructing: a Giant, Armed Nervous System', *Defense One* (26 September 2017) <http://www.defenseone.com/technology/2017/09/future-us-military-constructing-giant-armed-nervous-system/141303/> accessed 8 September 2018.

[145] "This push is too new, and still too developmental, to have attracted much concern from the public or from Capitol Hill. But that will change": ibid.

[146] This is a cloud computing contract, which will provide the tools to help fulfil the military's vision of a highly data-integrated armed nervous system, from the homefront through to the tactical edge. See US Department of Defense, 'Joint Enterprise Defense Infrastructure (JEDI)', *US Department of Defense Memo* (6 November 2017) <https://www.nextgov.com/media/gbc/docs/pdfs_edit/121217fk1ng.pdf> accessed 8 September 2018.

7.2.4 Conclusion on Proportionality

The principle of proportionality will clearly present LAWS-deploying forces with their most difficult LOAC compliance challenge. Despite being an objective standard with several administrable features, its overwhelmingly indeterminate nature and the potentially infinite combinations of (incommensurate) value conflicts make it near-impossible to administer with automatic processing alone. However, with full use of the Joint Targeting process and an appropriate system design that optimises task allocation between man and machine – and assuming durable communications superiority – there are clear possibilities for designing-in MHC, for the necessary input of controlled processing. Arguably, this will enable LAWS-deploying forces to apply the proportionality principle in both TSs and TLC, to at least the legally required standard.

7.3 Precautionary Measures

Distinction and proportionality can pose difficult compliance challenges for LAWS-deploying forces. Accordingly, a separate but overlapping set of precautionary rules reaffirms these two principles,[147] and it provides a more concrete set of obligations to enhance compliance with them. Much of these rules have been addressed indirectly throughout 6.5 and 7.2, and directly in several existing works.[148] The following will provide a selective evaluation of these along with a few additional applications, before arguing for a) a possible interplay between active and passive precautions in the context of LAWS deployments, b) a case for elevating (active) precautions to the status of a LOAC principle, and c) some potential applications thereof.

7.3.1 The General Obligation Under Article 57(1), AP I

Article 57(1), AP I, provides:

> In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects.[149]

[147] *AP I Commentary*, ¶ 2189.

[148] For example, Schmitt and Thurnher (n 71), 259-62; Ian S. Henderson, Patrick Keane and Josh Liddy, 'Remote and Autonomous Warfare Systems: Precautions in Attack and Individual Accountability' in Jens David Ohlin (ed.), *Research Handbook on Remote Warfare* (Edward Elgar, 2017); Jeffrey S. Thurnher, 'Feasible Precautions in Attack and Autonomous Weapons' in von Heinegg, Frau and Singer (eds.) (n 15).

[149] Article 57(1), AP I; *AMW Manual*, Rules 30 and 34; CIHL, Rule 15.

This general obligation will be elaborated upon in 7.3.5.2. For now, it is worth noting that it applies to broader "military operations" (not just 'attacks') and it entails a repeat obligation ("constant care"), which is operationalised in subsequent paragraphs.

7.3.2 Specific Treaty-Based Precautions Under Article 57(2) and (3)

Paragraphs (2) and (3) pertain to "attacks", which are defined as "acts of violence against the adversary, whether in offence or in defence".[150] Thus, being more narrowly framed than Paragraph (1) and – in the case of Paragraph (2)(a) – addressed specifically to "those who plan or decide upon an attack", the concrete rules apply to tactical commanders (who authorise deployments) and their battle staffs (who execute those orders).[151] That said, the obligations are technologically agnostic:[152] there is no stipulation that attacks must be planned or executed by people, just that responsible persons must discharge the listed obligations. This is even more clearly the case with Paragraphs (2)(b)-(c) and (3), which are expressed in the passive mood.[153]

7.3.2.1 Target Verification

Paragraph (2)(a)(i) requires these tactical staffs to:

> [D]o everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives…[154]

This distinction-based obligation has been extensively – albeit indirectly – covered throughout 6.5. The 'feasibility' constraint means that the obligation only applies to the extent that it is "practicable or practically possible, taking into account all circumstances prevailing at the time, including humanitarian and military considerations".[155] The first part of the definition of feasible makes clear that this is an obligation of *conduct*, not result.[156] Compliance depends on due diligence by way of

---

[150] Article 49(1), AP I. As the *AP I Commentary* makes clear at ¶ 2188, the following rules apply equally to aggressors and to victims of aggression.

[151] Ford (n 114), 448.

[152] Boothby (n 159), 41.

[153] Boothby (n 20), 120 (noting that passive provisions focus on the obligation, not the person or thing discharging it).

[154] Article 57(2)(a)(i), AP I; *AMW Manual*, Rules 31 and 32(a); CIHL, Rule 16.

[155] *AMW Manual*, Rule 1(q).

[156] Kimberley N. Trapp, 'A Framework of Analysis for Assessing Compliance of LAWS with IHL Precautionary Measures' in Robin Geiß (ed.), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016), 286.

verification *processes*, rather than outcomes.[157] The middle part makes clear that this is an *ex ante* obligation,[158] while the latter part of the definition clearly reflects the military necessity-humanity balance. Thus, if target verification is only possible using manned or remotely-piloted systems, the inability of a LAWS to perform to the same standard does not render this obligation infeasible, as an alternative means to achieve it is available.[159] On the other hand, issues of force protection are a 'military consideration',[160] which may counteract this and increase the likelihood of a given LAWS being deemed a feasible option.[161]

The target verification provision has four implications for LAWS. First, in TLC and at the point of weapons release in a TS, it will almost certainly require full use of onboard sensors, to recognise legitimate targets to the correct confidence threshold.[162] However, in some circumstances, the use of *external* sensors[163] – perhaps distributed via swarming[164] – or newer technologies like 'ghost imaging'[165] may also be required, where these are *available* and they have a *mitigating effect*;[166] and where it is *feasible* to expect their use.[167] A complementary point – in the case of aerial LAWS – concerns the altitude of deployment, which should be low enough (and confidence thresholds correspondingly high enough) to permit target verification[168] to at least human levels

[157] *OTP Report* (n 13), ¶ 29.

[158] *AMW Manual Commentary*, Rule 1(q), ¶ 4.

[159] William Boothby, 'Dehumanization: Is There a Legal Problem Under Article 36?' in von Heinegg, Frau and Singer (eds.) (n 15), 41. See also *AMW Manual Commentary*, Rule 39, ¶ 6.

[160] Dinstein (n 23), 168; *AMW Manual Commentary*, Rule 1(q), ¶ 5.

[161] *AMW Manual Commentary*, ibid., ¶ 6 (stating this to be a "matter of common sense and good faith").

[162] Schmitt and Thurnher (n 71), 260; *AMW Manual Commentary*, Rule 39, ¶ 4.

[163] *AP I Commentary*, ¶ 2195 (noting the example of launching aerial reconnaissance. Here, this may be done via surveillance drones transmitting additional sensory data directly to LAWS units).

[164] See 2.5.5.1, and n 288 therein.

[165] See 2.5.5.1, and n 290 therein.

[166] *AMW Manual Commentary*, Rule 1(q), ¶ 2.

[167] Dinstein (n 23) explains at 170 that mere *availability* of a mitigating technology, even with full *possibility* of its use, is not enough. Commanders have limited resources at their disposal and often plan multiple missions, where a subsequent use of a particular resource may have a greater *overall* mitigating effect on civilian risk.

[168] Daniel L. Haulman, 'The US Air Force in the Air War over Serbia, 1999' (2015) 62 Air Power History 6, 12 and 14 (noting the high altitude of NATO missions, which degraded the accuracy of air strikes because small targets like tanks were not easily discernible, and in one instance refugee convoys were hit when pilots confused long columns of tractors for tanks).

of accuracy.[169] To remain feasible, however, this may have to be considered alongside issues of preserving combat capability.[170]

Second, in the context of a TS planned under the deliberate targeting cycle, Article 57(2)(a)(i) puts relatively more emphasis on targeteers at Phase 2 (target development) and Phase 3 (capabilities analysis) to verify all aspects of target and weapon selection that rely on controlled processing. This leaves the LAWS to execute missions with the speed and precision of automatic processing in the field.[171] As explained in 5.5.3, a particularly vexing issue here is the use of decision-support systems such as *Project Maven*. These will need to be thoroughly tested, to ensure they enhance and do not undermine target verification by the staffs.

Third, recall the technologically agnostic nature of Article 57. In planning for either TLC or a TS, commanders are legally required to use whatever means will discharge the target verification obligation most effectively, so long as their use is feasible.[172] If a LAWS is able to verify targets more accurately than manned or remotely-piloted systems, because it uses a wider range of sensors and sources (including spectrums that are imperceptible to the human senses), and it processes these faster and without the fears and frailties of human imperfections,[173] then this may *obligate* the use of LAWS.[174]

Finally, recall from 6.5.2.1.1 that despite some calls for capture in lieu of lethal force, this is *not* a legal requirement under LOAC, which continues to authorise status-based targeting in armed conflict. Hence, to verify targets, an anti-personnel LAWS can arguably rely on the three-/four-part criteria for combatant status, and will not be required to undertake impractical conduct-based evaluations.[175]

---

[169] Sassóli (n 52), 320; *AMW Manual Commentary*, Rule 39, ¶ 4.

[170] Haulman (n 168), 12 and 13 (noting that high-altitude missions were necessary to avoid effective air defences).

[171] Hence, much depends on the robustness of the targeting process in matching specific TSs and their operational environment with the appropriate LAWS, or other capability. See 6.5.5.

[172] Dinstein (n 23), 170.

[173] Henderson, Keane and Liddy (n 148), 341-42.

[174] Sassóli (n 52), 320; Schmitt and Thurnher (n 71), 260; Weizmann (n 73), 16; *AMW Manual Commentary*, Rule 39, ¶ 6.

[175] See 6.5.2.1.1.

7.3.2.2 Minimise Collateral Damage

The obligation to minimise collateral damage is a precursor to the proportionality assessment, and it exists in two forms: choice of means and method, and choice of target. On the former, Paragraph (2)(a)(ii) provides that those planning and deciding on attacks must:

> [T]ake all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event minimizing, [collateral damage].[176]

Everything noted above in relation to feasibility applies equally to this provision. Consequently, a LAWS must not be deployed if other weapon systems are available that may inflict lower ECD, for a given MAA. The flipside is that if autonomous attack is likely to minimise collateral damage in the circumstances (for similar reasons discussed above) and if LAWS remain a feasible option, then Article 57(2)(a)(ii) will positively obligate their deployment and use.[177]

Aside from the existential question of 'means', Paragraph (2)(a)(ii) also concerns 'methods of attack'. Part of this goes back to commander decision-making during the targeting process. For example, in deciding the broader timing of attack (whether day or night) or by loading an alternative munition with a smaller blast radius, these may enable the attack to deliver the same military advantage but with significantly lower collateral effects.[178] Other human decisions that require controlled processing include weaponeering; the choice of aerial or ground-based LAWS; in the case of aerial LAWS, setting upper and lower boundaries for the altitude of attack; and whether advance warnings should be issued. These and other considerations mostly take place during Phases 3 (capabilities analysis) and 5a (mission planning).[179] Moreover, a LAWS on deployment with integrated CDEM may utilise its automatic processing capabilities to minimise collateral damage. For example:

---

[176] Article 57(2)(a)(ii), AP I; CIHL, Rule 17.

[177] Sassóli (n 52), 320; Schmitt and Thurnher (n 71), 261-62; Thurnher (n 148), 114 (arguing that this highlights the unintended consequences of a LAWS ban, as it would deprive commanders of a potentially more humane option).

[178] Both examples from *AP I Commentary*, ¶ 2200.

[179] See 5.3.1.3 and 5.3.1.5.

- By quantifying the ECD of various onboard weapons and directions of attack, it can select the exact munition and approach angle that will achieve its mission goal with the lowest collateral damage.[180]
- Unlike a 'fire-and-forget' munition, a LAWS can loiter until passing civilians have moved out of the collateral effects radius;[181] or it can follow a moving target into an isolated or less populated area, before weapons release.[182]
- Within the altitude boundaries set by the commander, an aerial LAWS may calculate an optimum altitude in the circumstances, to minimise a weighted risk factor.
- A LAWS may issue its own advance warnings, to clear the collateral effects radius.[183]

Paragraph (3), which complements the above, provides:

> When a choice is possible between several military objectives for obtaining a similar military advantage, the objective to be selected shall be that the attack on which may be expected to cause the least [civilian risk].[184]

An example may be an attack on a railway line, which aims to block a vital supply route, and where the same MAA is gained from attacking close to a highly populated train station compared with a more remote and uninhabited area.[185] Either way, the enemy's supply route is cut off, though in the latter scenario there is little, if any collateral damage. This task would seem to be amenable to automatic processing, as a LAWS can simply be programmed to recognise that attacking any point along a rail (or road) network is equally advantageous; beyond that, its integrated CDEM would enable it to select the part of the objective that poses the lowest civilian risk.[186]

---

[180] Thurnher (n 148), 113.

[181] Boothby (n 73), 118. See also 7.2.3.3.

[182] Thurnher (n 148), 113.

[183] See 7.3.2.4.

[184] Article 57(3), AP I; *AMW Manual*, Rule 33; CIHL, Rule 21.

[185] *AP I Commentary*, ¶ 2227.

[186] As noted in the *AMW Manual Commentary*, Rule 33, ¶ 5, similar reasoning applies to a power-generating facility and its various transformers, substations and power transmission lines. Where the objective is merely to disrupt power supplies to enemy forces, attacking any of these will achieve that goal, but targeting one of the latter three will do so with relatively less civilian risk. Even within those three categories, it is likely that integrated CDEM will enable a LAWS to select the one posing the objectively lowest collateral risk, so long as the equivalence of the MAA is pre-programmed.

However, in many cases target choices are less clear-cut, and this may necessitate complex contextual judgments to assess the closeness of their MAA values.[187] Moreover, lack of certainty often brings in an element of subjectivity in these assessments,[188] and this extends to the composition of ECD.[189] In such instances, the choice of objective will likely be a matter for commanders and their battle staffs through one of the Joint Targeting Cycles, with the LAWS simply executing a TS. Even then, it has been argued that the task of assigning particular MAA values to alternative but dissimilar targets is "next to impossible", such that Article 57(3) will "only occasionally be applicable in practice".[190]

7.3.2.3 Cancel or Suspend Attacks

Paragraph (2)(b) provides that:

> [A]n attack shall be cancelled or suspended if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack may be expected to [violate the proportionality rule].[191]

This provision applies not only to those who plan or decide upon an attack, but also and *primarily* to those who execute it.[192] Arguably, this puts relatively more onus on the LAWS sensors and control software, given the passive mood of the provision and that the counterfactual is often a pilot or weapons operator in the field.[193] Nonetheless, Paragraph (2)(b) is only engaged where the likely violation becomes "apparent". In most modern TSs, this is unlikely as front-line attackers – particularly at standoff ranges – rarely see in advance what they attack.[194] Moreover, pilots are required to act on information in their target package[195] and not to question it; especially when they possess so little contextual knowledge of the overall operation compared with senior

---

[187] Thurnher (n 148), 112-13 (referring to the friendly force risk posed by different targets, and the holistic nature of MAA).

[188] Dinstein (n 23), 166-67.

[189] Henderson (n 25), 190-92 (questioning how many homes are worth a civilian life, in the context of choosing from several targets of equivalent MAA).

[190] Ibid., 189.

[191] Article 57(2)(b), AP I; *AMW Manual*, Rule 35; CIHL, Rule 19.

[192] *AP I Commentary*, ¶ 2220. In this sense, it exemplifies the constant care obligation in Paragraph (1).

[193] Ibid., ¶ 2221; *AMW Manual Commentary*, Rule 35, ¶ 3.

[194] Hans Blix, 'Means and Methods of Combat' in UNESCO, *International Dimensions of Humanitarian Law* (Martinus Nijhoff, 1988), 147.

[195] See n 256 in Chapter 6, on the typical contents of the target package given to F-16 pilots.

commanders.[196] Legally, they are entitled to rely on instructions and, where images of the target and its surrounding area are provided, they may accept these as fully verified by the commander and his battle staffs.[197] Front-line attackers should therefore only cancel or suspend an attack if it becomes apparent that something *unbriefed in the target package* – be that a protected object[198] or an unexpected mass of civilians[199] – appears without prior warning.

Accordingly, for a LAWS deployment to comply with Article 57(2)(b), it need only be programmed to cancel or suspend an attack in the event that the target or target area no longer meets its programmed parameters to the correct confidence threshold.[200] Examples of how this might occur were provided earlier, in relation to both distinction[201] and proportionality.[202] Whether a *specific* weapon system complies with this rule will depend on whether it recognises the situation that calls for cancelation/suspension, and thereby refrains from launching an attack *at least to the standard of a human operator in the same circumstances*.[203]

Certainly, in the right operational environment, the precision and responsiveness of autonomous control systems may even exceed human performance.[204] This is because of the relatively short response time after a target is acquired and its vicinity reconnoitred, which tends to lower the risk of situational changes in which civilians unexpectedly enter the collateral effects area.[205] Moreover, in the event that civilians

---

[196] See (notes and text accompanying) nn 30-37.

[197] *OTP Report* (n 13), ¶ 85.

[198] Paul Scharre, 'Centaur Warfighting: The False Choice of Humans Vs. Automation', (2016) 30 Temple International & Comparative Law Journal 151, 154 (describing an incident during the Kosovo air campaign, where an F-15E Weapon Systems Officer (WSO) launched a remotely-flown standoff weapon at an alleged radar site. Twelve seconds before impact, he noticed an unbriefed profile of a church steeple, prompting him to steer the weapon into an empty field).

[199] Henderson (n 25), 184-85 (noting that Australian F/A-18 pilots occasionally pulled out of bombing raids at the last minute when they detected an unbriefed civilian presence, presumably to request a proportionality reassessment).

[200] Henderson, Keane and Liddy (n 148), 355.

[201] See, for example, 6.5.2.2 on sparing human visual and heat signatures that do not conform to the three-/four-part criteria for active combatant status; 6.5.3.4.1 on recognising the distinctive architecture of churches and mosques, similar to the F-15E WSO, above; and 6.5.3.4.5 on detecting distinctive signs and emblems. In all these cases, a LAWS would be programmed to hold fire.

[202] See 7.2.3.3 on possible ways to cancel or suspend a potentially 'disproportionate' attack.

[203] Sassóli (n 52), 320; *AMW Manual Commentary*, Rule 39, ¶ 4.

[204] Wolfgang Richter, 'Military Rationale for Autonomous Functions in Weapons Systems', *Presentation at the 2015 Meeting of Experts on LAWS* (13-17 April 2015), 4 (on file with author).

[205] Ibid. (noting that while standoff-air-to-ground weapons and cruise missiles must fly several minutes, or even hours before hitting the identified target, LAWS will be able to do this within seconds).

do unexpectedly enter the target area, a LAWS will arguably be faster than humans to recognise this,[206] and to assess the risk of further movement into the collateral effects zone. In such a case, the system may be better able to cancel or suspend the attack before any civilian harm occurs. The experiences of NATO forces in Grdelica, April 1999, are particularly instructive.[207] Here, the Weapon Systems Officer of an F-15E *Strike Eagle* made two recognition failures, resulting in a double-bombing of a planned target (a railway bridge), just as a passenger train appeared and moved further into the target area; this resulted in ten deaths and 15 injuries.[208] The *OTP Report* noted the difficulty of multi-tasking in a high-speed jet, the short reaction time of seven or eight seconds, and the potential human error.[209] Arguably, a LAWS – unencumbered by GPS-latency, G-lock, or neuromuscular delay – would have avoided this, by virtue of its superior sensory perception, its automatic processing capabilities, and the absence of any biological limitations.

7.3.2.4 Effective Advance Warning

Finally, "effective advance warning" must be given to civilians that may be affected by an impending attack,[210] unless circumstances do not permit.[211] The aim is that as many civilians as possible leave the area, or at least get out of the collateral effects zone. In a LAWS context, commanders will have to determine during the targeting process whether advance warning is compatible with mission goals; and if so, whether this should be done through traditional means (e.g. broadcast messages, or pamphlet-drops),[212] or via the LAWS. If they opt for the latter, integrated CDEM would likely enable the detection of civilians and civilian objects, hence the determination of whether a warning is needed in the particular circumstances.[213] If so, a platform fitted with non-lethal munitions can discharge these before weapons release,[214] similar to the

---

[206] For example, via the proxy of 'human heat signatures not positively identified as combatants'. A LAWS will process these data and act instantly, while humans exhibit a neuromuscular delay of around 0.25 seconds.

[207] See *OTP Report* (n 13), ¶¶ 58-62.

[208] Ibid., ¶ 58.

[209] Ibid., ¶¶ 61-62.

[210] Article 57(2)(c), AP I; *AMW Manual*, Rule 37; CIHL, Rule 20.

[211] Again, feasibility is an important precondition. Thus, if an element of surprise is necessary for the attack to deliver a military advantage, then the obligation to provide a warning may not apply: *AP I Commentary*, ¶ 2223; *AMW Manual Commentary*, Rule 37, ¶ 4.

[212] *AP I Commentary*, ¶ 2224.

[213] Because Article 57(2)(c) only requires advance warning of attacks "which may affect the civilian population".

[214] *AMW Manual Commentary*, Rule 37, ¶ 11 (noting the possibility of firing warning shots).

Israeli 'roof-knocking' technique.[215] Once a non-lethal munition is released, the LAWS may then loiter and survey the area, where it performs repeat split-second calculations on fleeing civilians and vehicles in relation to the collateral effects radius. At the optimum time of attack, when observable civilians are sufficiently out of the kill zone, the weapon system can then discharge its lethal munitions.[216]

7.3.3 The 'Obligation' to Take Passive Precautions Under Article 58

Article 58, AP I, concerns *passive* precautions, or 'precautions *against the effects* of attack', which are to be taken by the Party in control of territory, not necessarily the one launching an attack.

Specifically, Article 58 requires "to the maximum extent feasible" that the Party exercising control over civilians and civilian objects:

a) "endeavour to remove" these from the vicinity of military objectives;[217] and
b) avoid locating military objectives within, or near densely populated areas.[218]

This dual-obligation to unclutter the battlefield clearly works to enhance LAWS operations, by requiring the maintenance of a simpler operational environment. In turn, this will make the battlefield relatively more amenable to automatic target recognition (ATR). On the other hand, the 'feasibility' qualification and other wording ("endeavour") negates the idea that Article 58 is a strict obligation, some commentators tending to view it more as a 'recommendation'.[219]

More generally, there is an open-ended obligation to take "other necessary precautions" to protect civilians and their objects from the dangers of military operations.[220] This may include the provision of shelters and well-trained civil defence

---

[215] 'Roof-knocking' is where an aircraft "targets a building with a loud but non-lethal bomb that warns civilians that they are in the vicinity of a weapons cache or other target", thereby allowing "all residents to leave the area before the [Israeli Defense Force] targets the site with live ammunition". See Israeli Defense Forces, 'How is the IDF Minimizing Harm to Civilians in Gaza?' *IDF Blog* (16 July 2014) <https://www.idf.il/en/minisites/hamas/how-is-the-idf-minimizing-harm-to-civilians-in-gaza/> accessed 8 September 2018.

[216] Alternatively, a differentiated swarm consisting of lethal and non-lethal micro-drones can do the same, while a standalone munition may fly low over the target before dive-bombing into it: *AP I Commentary*, ¶ 2224.

[217] Article 58(a), AP I; *AMW Manual*, Rule 43; CIHL, Rule 24.

[218] Article 58(b), AP I; *AMW Manual*, Rule 42; CIHL, Rule 23.

[219] Dinstein (n 23), 173.

[220] Article 58(c), AP I; *AMW Manual*, Rule 44; CIHL, Rule 22.

services;[221] air-warning systems and air-raid shelters;[222] and the deployment of protective emblems, signs and signals to assist discriminatory targeting.[223] The *AMW Manual Commentary* specifically extends this obligation to attacks involving unmanned combat aerial vehicles (UCAVs),[224] and it acknowledges that in modern air and missile warfare it may be "necessary to consider other methods than marking in order to bring protected locations to the notice of the enemy".[225]

7.3.4 A Potential Interplay Between Articles 57 and 58

Importantly, there may be an interplay between Articles 57 and 58; namely, between the rules and obligations to take active and passive precautions. For at a more fundamental level, advance warnings under Article 57(2)(c) do not have to relate to specific or impending attacks, but "may also have a general character".[226] Thus, publicly announcing that autonomous systems relying on ATR will be deployed, but without necessarily specifying in which precise attacks, may enhance the obligation of the Party in control of civilians and civilian objects to take relevant passive precautions under Article 58(c).[227] This may include the following:

- Ensuring that protective emblems, signs and signals are deployed to assist discriminatory targeting by technical means.[228]
- Ensuring these are produced with materials and paints that facilitate detection by ATR.
- Providing GPS coordinates to help attacking forces avoid protected objects in fixed locations, which may be difficult to detect solely via ATR.

---

[221] *AP I Commentary*, ¶¶ 2257-2258.

[222] *AMW Manual Commentary*, Rule 44, ¶ 2.

[223] See 6.5.3.4.5.

[224] *AMW Manual Commentary*, Rule 44, ¶ 3.

[225] Ibid., Rule 42, ¶ 3. An obvious method would be to use materials that make signs and emblems recognisable by technical means of detection. Another may be to provide GPS coordinates of all protected fixed objects.

[226] *AP I Commentary*, ¶ 2225 (noting the possibility to warn that "certain types of installations or factories" may be attacked, or even containing "a list of the objectives that will be attacked").

[227] This is an extension to *AMW Manual Commentary*, Rule 44, ¶ 4, which articulates a similar approach in relation to UCAVs because of their smaller visual, radar and noise signatures. These characteristics may allow greater penetrability into areas where air defences would normally deter manned aircraft attacks, hence the enhanced obligation of the Party in control of civilians and civilian objects to take passive precautions in those areas.

[228] See 6.5.3.4.5.

As above, this is all subject to the feasibility requirement. Thus, to effectively enhance the Article 58 obligation, advance warnings must themselves be 'effective' in the circumstances.[229] For example, it is likely that such warnings will have to be as geographically bounded as possible, and sufficiently far in advance of an attack to afford the Party in control of territory a reasonable opportunity to distribute the protective emblems and signs. Where this is likely to be problematic, it is possible that the LAWS-deploying side may also notify the International Committee of the Red Cross and interested non-governmental organisations, which may also assist in the sourcing and distribution of specially designed protective emblems and signs.

As explained in 6.5.3.4.6, there are legal protections in Articles 37, 38 and 85(3) that prohibit any improper use of the above safeguards. Despite these, however, the same caveat put forth in 6.5.3.4.5 applies here: attacking forces must remain independently vigilant and avoid any over-reliance on emblems and signs. Concretely, the safeguards are *in addition* to active precautions and other commander-led targeting efforts, and they do not permit or justify a watering down of such efforts.[230]

7.3.5 'Elevating' Precautions to the Status of a LOAC Principle

7.3.5.1 Precautions as the Lynchpin of the Targeting Process

One of the most striking features of the above precautionary rules is their relatively concrete and objective nature. This is not absolutely the case, as the feasibility requirement and the need to assess MAA in some instances calls for complex value judgments. However, compared with the amorphous proportionality principle, the Article 57 rules do afford the commander a strong package of measures for civilian risk mitigation; an effective counter-weight to military necessity, yet one that is well-aligned with military operational and tactical logic.[231] In some instances, the precautions will entail lowering ECD through smarter planning;[232] in other instances,

[229] Boothby (n 20), 127-29.

[230] Article 51(8), AP I; *AMW Manual Commentary*, Rules 46 and 37, ¶ 16.

[231] Geoffrey S. Corn, 'War, Law, and the Oft Overlooked Value of Process as a Precautionary Measure' (2015) 42 Pepperdine Law Review 419, 423 (arguing that commanders often see feasible precautions as "more rationally aligned with the overall tactical and operational concept of operations").

[232] For example, through the choice of target, weaponeering, or varying the time of deployment.

they entail objectively ascertainable tasks or determinations that are amenable to automatic processing.[233]

In the chronology of targeting presented in 5.3, addressing the distinction principle is clearly the first step in civilian risk mitigation. Next, the commander and his battle staffs consider an array of precautionary measures, before applying the proportionality principle as a final protective layer for civilians.[234]

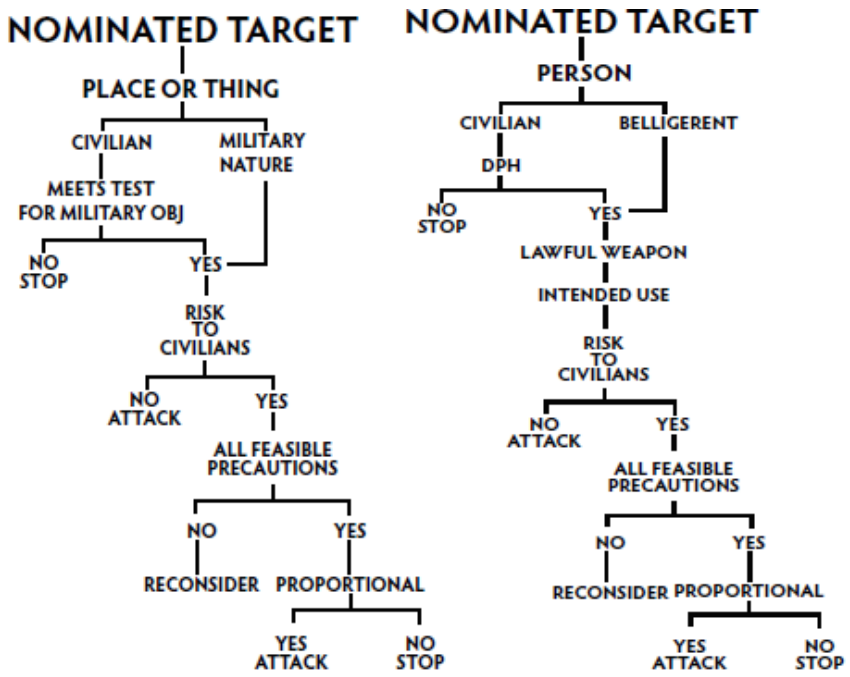

Figure 7.1: The chronology of targeting. Source: Corn (n 231), 436.

Accordingly, precautionary measures "bridge the conceptual borderline between distinction and proportionality"[235] (see Figure 7.1, above), such that they mitigate the difficulty of applying the proportionality principle as a final step. This they do by minimising the ECD in every feasible way possible so that the proportionality balance will tip decisively in favour of the MAA, thereby obviating the need to make such complex judgment calls on the battlefield.[236] This approach is now officially

[233] For example, when assessing the area of a target with the lowest ECD, or in cancelling an attack that does not meet confidence thresholds.

[234] This accords with Dill's three-part criteria for operationalising proportionality: adequacy, necessity, and commensurability (at n 65).

[235] Geoffrey S. Corn and James A. Schoettler, 'Targeting and Civilian Risk Mitigation: The Essential Role of Precautionary Measures' (2015) 223 Military Law Review 785, 799.

[236] Ibid.; Corn (n 231); Kenneth Watkin, 'Military Advantage: A Matter of Value, Strategy and Tactics' (2014) 17 Yearbook of International Humanitarian Law 277, 311-14.

documented in the US *DoD Manual*,[237] which notes that the requirement to take feasible precautions and the proportionality principle are "mutually reinforcing obligations".[238] As LAWS will not be able to undertake their own proportionality assessments, this clearly has value for commanders contemplating such deployments. Namely, well-planned precautions will increase the likelihood that an autonomous attack will remain within the "zone of proportionality",[239] and indeed this is exactly the approach taken by Thurnher and Boothby, above.[240] However, not all the precautionary rules will apply equally (or at all) in every LAWS deployment; while in other deployments some hitherto unfamiliar precautions may be beneficial. This raises the question as to whether 'elevating' precautions to the status of a full LOAC principle may positively assist future LAWS-deploying forces.

7.3.5.2 A Precautionary Principle?

Recall from 3.2.2.2.1 that there is a fundamental difference between rules and standards/principles, based on whether the exact content of the norm is determined before (rules) or after (standards/principles) relevant facts have materialised.[241] Consequently, rules are relatively precise and applicable only to the specific contexts anticipated in their terms, while principles are more vaguely drafted and they serve as a general source of guidance.[242] As a corollary, principles guide the interpretation and application of specific (under-inclusive) rules, and they guide decision-making *where no discernible rule exists*.[243] Thus, while both rules and principles enjoy the same (binding) normative status, they differ greatly in their scope and breadth of application.

Distinction and proportionality are unquestionably seen as "cardinal principles" of IHL/LOAC,[244] and they command most legal and public attention during an armed conflict.[245] By contrast, the precautionary measures tend to be under-valued in IHL discourse, and this may reflect the understanding of these measures as rules-based

---

[237] US Department of Defense, *Law of War Manual* (DoD, 2015; December 2016 Update), § 5.10.
[238] Ibid., § 5.10.5.
[239] *Targeted Killings Case*, ¶ 58.
[240] See 7.2.3.2 and 7.2.3.3.
[241] Louis Kaplow, 'Rules Versus Standards: An Economic Analysis' (1992) 42 Duke Law Journal 557.
[242] Ibid.
[243] Corn and Schoettler (n 235), 833.
[244] *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion) [1996] ICJ Reports 226, ¶ 78.
[245] Watkin (n 236), 311.

obligations.[246] Moreover, being restricted to 'attacks', the rules are not instinctively integrated into other aspects of military operations, such as training and staffing.[247]

However, this approach is arguably wrong as it fails to account for the pervasive obligation in Article 57(1), AP I. The lynchpin of this provision is to take "constant care" in the conduct of "military operations" *vis-à-vis* civilian protection.[248] Aside from the *AP I Commentary* specifically describing this as a "general principle",[249] there are two other reasons why this is arguably true. First, 'constant care' – while not defined – clearly imposes a repeat obligation; thus, it is insufficient to take care during pre-deployment preparations but to then ignore changes in civilian risk once a LAWS is in-flight, or vice versa.[250] Second, this obligation should inform all aspect of 'military operations', not just 'attacks'. While the latter are merely "acts of violence against the adversary…",[251] military operations are more broadly defined as "any movements, manoeuvres and other activities whatsoever carried out by the armed forces with a view to combat".[252] As noted above, this may include training and staffing issues, amongst others.

Accordingly, 'constant care' is a pervasive obligation, which is incumbent upon all persons who have control over the use and deployment of LAWS.[253] Paragraphs (2) and (3) reflect this broader principle and are merely practical applications of it,[254] but they should not be seen as the limit of that principle. Namely, the specific rules are illustrative, but not exhaustive.

In a US/NATO context, the Joint Targeting process arguably provides a basis for fully implementing precautions as a principle. This is because of the symmetry between operational practice and targeting law, and the elaborate nature of the targeting process itself. Together, these enable commanders and their staffs to integrate precautionary

[246] Corn and Schoettler (n 235), 834.
[247] Ibid., 835.
[248] See also CIHL, Rule 15; *AMW Manual*, Rules 30 and 34.
[249] *AP I Commentary*, ¶ 2191.
[250] Ford (n 114), 448.
[251] Article 49(1), AP I.
[252] *AP I Commentary*, ¶ 2191.
[253] Boothby (n 159), 41.
[254] *AP I Commentary*, ¶ 2191.

measures *earlier on* during the attack-planning process and to a *greater extent* than the legal minimum, such that the outcome is to increase the range of precautionary options that come to light, and to increase the likelihood of their uptake.[255] For LAWS-deploying forces, which may find that the listed precautions do not always suffice, this provides a useful starting point to ensure that precautionary measures are broadly-conceived, pervasive and fit for purpose in a LAWS context.

7.3.6 Potential Applications of a Precautions Principle for LAWS Deployments

More specifically, elevating precautions to the status of a full LOAC principle will arguably assist in reducing the margin of error in LAWS operations, thereby bringing these more comfortably within the commander's margin of appreciation. This will occur as specific and apt precautionary measures will enable more predictable operation in the field, and thus attacks may be expected to comply more effectively with the principles of distinction and proportionality.

The following examples illustrate how, both within and beyond the Joint Targeting process, more LAWS-specific precautionary measures may arise. In some instances, these will draw insight from relevant specific weapons treaties, and in all cases the aim is to afford a degree of MHC to LAWS operations.

7.3.6.1 Front-Loading

Already implicit in much of the above, Lewis argues that a crucial precautionary measure is to undertake as many critical tasks as possible in a pre-deployment setting, thereby reducing the number of algorithmic decisions to be made on a chaotic battlefield.[256] In a LAWS context, such front-loading of critical tasks is likely to focus on those that require controlled processing, which will be undertaken by the commander and his battle staffs. Examples include the determination of the MAA of a HVT;[257] of a set of MAAs of alternative targets;[258] or the assessment of whether a dual-use object is being put to a military use or purpose.[259]

---

[255] Corn (n 231).

[256] Larry Lewis, 'Redefining Human Control: Lessons From the Battlefield for Autonomous Weapons', *Center for Autonomy and AI* (March 2018), 11-12 <https://www.cna.org/CNA_files/PDF/DOP-2018-U-017258-Final.pdf> accessed 10 June 2018.

[257] See 7.2.3.2.

[258] See 7.3.2.2.

[259] See 6.5.3.2.

In all cases, there is a 'division of labour' between man and machine, with the latter completing all aspects of the mission that require automatic processing. The challenge for future LAWS-deploying forces is therefore to separate out all tasks and decisions that require controlled processing, and to undertake these either in a pre-deployment setting, or on the battlefield using a dial-in capability.[260]

7.3.6.2 Developing Capabilities

Front-loading of critical tasks is undoubtedly enhanced when a State possesses the appropriate capabilities. In this connexion, recall that precautionary measures are obligations of conduct, which implicate due diligence.[261] In turn, this carries an obligation to *develop* relevant capabilities, which itself varies with a State's socio-political circumstances, its general technological capabilities, and its particular development of battlefield technologies.[262] Thus, a well-resourced State that is constantly engaged in (or threatened with) war may be expected to develop advanced surveillance and target identification capabilities; even outside of armed conflict, in preparation for it. Arguably, this due diligence obligation is not limited to technology, but extends to more painstaking precautionary activities. An example may be working with archaeologists to build a database of cultural property in a particular conflict zone, to enhance compliance with the customary rules restated in CIHL, Rule 38(A).[263]

7.3.6.3 A Precautionary Approach to Online Learning

Should machine learning take place 'online' (in the battlefield), there is the risk that a LAWS may 'learn the wrong lessons' and behave unpredictably, and this was vividly illustrated in 2.5.3.2. The most precautionary approach would be to restrict learning to 'offline' training exercises and laboratory scenarios, to ensure that any acquired unlawful behaviour occurs in a safe and non-critical context, and can be overridden by human programmers. The system would then be 'frozen' on deployment so that it cannot continue to learn new tasks, tactics or target sets. However, this may mean forgoing certain military advantages that States are unwilling to abandon;[264] for

---

[260] Similar to 7.2.3.4.

[261] Trapp (n 156), 286.

[262] Ibid., 288.

[263] See 6.5.3.4.1.

[264] James Farrant and Christopher M. Ford, 'Autonomous Weapons and Weapons Reviews: The UK Second International Weapon Review Forum' (2017) 93 International Law Studies 389, 406.

example, adapting to slight changes in the operational environment for greater targeting accuracy and precision.[265]

Accordingly, deploying forces may develop and implement 'procedural safeguards'.[266] An example would be to make algorithmic updates on the battlefield subject to human approval, perhaps in between sorties, before going 'live'.[267] Of course, this assumes the proposed updates are scrutable and intuitive; for example, through 'Explainable AI', so that the human reviewer can make an informed and accountable decision.[268] Also, depending on the nature and extent of the algorithmic update, it may obligate a whole new legal review under Article 36, AP I,[269] potentially requiring a legal review team to be despatched in the field.

7.3.6.4 Spatio-Temporal Restrictions[270]

Fourth, LAWS deployments may be subject to relatively tighter spatio-temporal restrictions, similar to other weapons that operate with humans out-of-the-narrow-loop. For example, Amended Protocol II to the Convention on Certain Conventional Weapons[271] requires human supervision of anti-personnel mines *within a confined, monitored and protected area*;[272] or, for mines emplaced outside such an area, to self-destruct or self-deactivate *within a maximum time period* as specified in the Technical Annex.[273] The Protocol then adopts a graduated approach, as it relaxes the spatial

---

[265] See 2.5.1.4 on reasons for online learning.

[266] Kush R. Varshney and Homa Alemzadeh, 'On the Safety of Machine Learning: Cyber-Physical Systems, Decision Sciences, and Data Products' (v2 22 August 2017) <https://arxiv.org/pdf/1610.01256v2.pdf> accessed 7 July 2018.

[267] Paul Christiano, 'Approval-Directed Algorithmic Learning', *AI Alignment Blog* (21 February 2016) <https://ai-alignment.com/approval-directed-algorithm-learning-bf1f8fad42cd> accessed 7 July 2018.

[268] See 2.5.3.3. Note that Explainable AI is still in development.

[269] Farrant and Ford (n 264), 406.

[270] For a fuller analysis of spatio-temporal restrictions in the context of extracting legal transplants for a LAWS regulatory regime, see Maziar Homayounnejad, 'Ensuring Fully Autonomous Weapons Systems Comply with the Rule of Distinction in Attack' in Stuart Casey-Maslen et al. (eds.), *Drones and Other Unmanned Weapons Systems under International Law* (Brill Nijhoff, 2018), 148-50.

[271] Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices (adopted 10 October 1980, amended 3 May 1996, entered into force 3 December 1998) 2048 UNTS 93.

[272] Article 5(2)(a), Amended Protocol II, lays down the triple requirement of durable and visible perimeter-marking; monitoring by military personnel; and fencing or other protection, to exclude civilians from the area.

[273] Sub-Paragraph 3(a) of which requires at least 90% of activated mines to self-destruct *within 30 days* of emplacement. The provision also requires a *back-up* self-deactivation feature, such that at least 99.9% of activated mines either self-destruct or self-deactivate *within 120 days* after emplacement.

restrictions for a maximum period of 72 hours, so long as certain additional requirements are met.[274]

These and other rules in Amended Protocol II aim to keep civilians out of the kill zone, while anti-personnel mines are operational and liable to cause them harm; and they limit the time during which there can be a non-human controlled threat to the civilian population. Accordingly, such rules bound the independent 'operation' of landmines in time and space, to ensure that accountable humans remain in control. Similar precautions may be taken by LAWS-deploying forces, so human operators can exercise frequent, or at least periodic control over potentially lethal engagements by robotic weapons.[275] Their controlled processing will be more likely to recognise changing battlefield circumstances that a LAWS will not perceive, thereby enabling operators to adjust deployments in a timely manner. As a corollary, tighter spatio-temporal restrictions will also safeguard the opportunity to shift to a law enforcement model (as changing circumstances dictate), wherein human rights norms will play a greater role.[276]

There is no way to specify an appropriate set of spatio-temporal limits in the abstract; rather, these will likely emerge from prior testing and evaluation of specific LAWS models, and they should also be subject to adjustment by commanders in a given deployment scenario. In doing so, commanders may also wish to consider the 'boxed autonomy' approach,[277] which combines the three-dimensional 'kill-box' with tight target parameters (see below) and exhaustive front-loading.[278]

7.3.6.5 Target Parameters[279]

The landmines regime also contains several provisions that *effectively* set target parameters in order to retain advance human control over the kinds of persons or

---

[274] Article 5(6), Amended Protocol II. The additional requirements are a) the munitions propels fragments in a horizontal arc of less than 90 degrees; and b) they are placed on or above (not below) the ground.

[275] Maya Brehm, 'Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems Under International Humanitarian and Human Rights Law', *Geneva Academy Research Brief* (2017).

[276] Ibid. Namely, where there will be a preference for non-lethal options, such as capture and arrest.

[277] International Panel on the Regulation of Autonomous Weapons (iPRAW), 'Focus on Technology and Application of Autonomous Weapons, *"Focus on" Report No. 1* (August 2017), 15-16.

[278] See also Homayounnejad (n 270), 139 (discussing kill-boxes as a possible application of the landmines regime to LAWS).

[279] For a fuller analysis of LAWS, target parameters and the landmines regime, see ibid., 150-52.

objects that are targeted. Thus, Article 1 of the Mine Ban Treaty[280] comprehensively prohibits "anti-personnel mines", which are defined as devices that are designed to target a "person", while excluding from its purview anti-tank mines.[281] Similarly, Article 2(3), Amended Protocol II, states that its more permissive, yet highly precautionary approach[282] only applies to mines that are "primarily designed to be exploded by the presence, proximity or contact of a *person*…"[283] By contrast, Article 1(1) specifically excludes any application of the Protocol to anti-ship mines at sea, where there is unlikely to be a concentration of civilians; assuming these are not deployed along commercial shipping routes.

Again, these provisions aim to keep civilians out of harm's way, but this time by bounding the *kinds* of targets against which mines may be triggered. By imposing relatively permissive rules on anti-tank mines (with more resistant pressure plates)[284] or anti-ship mines (designed for deployment at sea or in inland waterways), we see lighter regulation on those munitions whose effective target parameters are limited to heavy, durable military objects. While persons may be killed during these mine deployments, this is a secondary consequence of targeting the military object, and it is assumed that the inhabitants of a tank or a ship sailing along certain routes are, in any event, likely to be targetable combatants.[285] By contrast, anti-personnel mines, which have relatively broader target parameters via the proxy of more sensitive pressure plates or tripwires,[286] carry a greater risk of unintended engagements with civilians, including children. This explains the correspondingly greater precautionary requirements.

---

[280] Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction (adopted 18 September 1997, entered into force 1 March 1999) 2056 UNTS 241.

[281] Article 2(1), Mine Ban Treaty (specifically excluding "[m]ines designed to be detonated by…a *vehicle* as opposed to a person") (emphasis added).

[282] Several key military States, such as China, Russia and the US are not Party to the Mine Ban Treaty, on the grounds of military necessity. Instead, they have signed up to the permissive yet precautionary Amended Protocol II.

[283] Article 2(3), Amended Protocol II (emphasis added).

[284] Most anti-tank mines require a weight force of mass in excess of 100 kilograms for detonation.

[285] Though this assumption is not perfect, and it is not inconceivable that a warship might contain medical personnel and/or chaplains, who are not targetable combatants under Article 43(2), AP I.

[286] While thresholds differ, they can be as low as 5-9 kilograms of weight force, depending on the munition.

This graduated approach, which links the breadth of target parameters to the normative constraints on the commander, may provide inspiration to LAWS-deploying forces too. Namely, there is a risk that systems programmed with broad parameters ('large vehicle', with a given set of dimensions) and deployed in areas where civilian and military objects comingle may inadvertently engage civilian objects, like trucks or buses. In such a case, it would be prudent to narrow the target parameters as much as possible; for example, commanders may specify exact models that are both *exclusive* to the enemy and *recognisable* to the ATR systems, like '*T-80* tank'.[287] In line with the graduated approach, this degree of specificity may be relaxed in simpler operational environments; for example, an undersea LAWS programmed to pursue 'large metallic objects with an acoustic signature' may be lawful without any further precautions, as there is unlikely to be any civilian objects fitting those parameters in that particular context. However, it is arguably not possible to lay down *ex ante* bright-lines. Instead, 'rules of thumb' would likely be more administrable, such as requiring commanders to opt for the narrowest possible target parameters, consistent with ATR capabilities and the exigencies of military necessity in the operational environment.[288]

7.3.6.6 Upper Engagement Limits[289]

To complement the above three constraints, upper engagement limits (UEL) may also offer a precautionary measure that enhances human control. This is especially useful where a LAWS is deployed for TLC, to attack numerous specific targets in a number of different target categories; and/or where there is a high degree of comingling, with a correspondingly greater civilian risk. In this regard, the Convention on Cluster Munitions[290] (CCM) may offer some useful pointers within its Article 2(2)(c) technical criteria.[291] Specifically, criterion (i) requires fewer than ten explosive submunitions

---

[287] See 4.5.1.

[288] Another possible application of the landmines regime is to restrict LAWS to anti-material targeting; or even Canning's approach of targeting either the 'bow' or the 'arrow' but not the 'human archer'. See 6.5.5.

[289] For a fuller analysis of upper engagement limits applied to LAWS in TLC and TSs, respectively, see Maziar Homayounnejad, 'Ensuring Lethal Autonomous Weapon Systems Comply with International Humanitarian Law, *TLI Think! Paper 85/2017* (2017), 55-58 <https://ssrn.com/abstract=3073893>; Maziar Homayounnejad, 'The Lawful Use of Autonomous Weapon Systems for Targeted Strikes (Part 2): Targeting Law & Practice', *TLI Think! Paper 13/2018* (2018), 62-68 <https://ssrn.com/abstract=3200416> both accessed 2 October 2018.

[290] Convention on Cluster Munitions (adopted 30 May 2008, entered into force 1 August 2010) 2688 UNTS 190.

[291] More precisely, for a weapon to be excluded from the definition of the (prohibited) 'cluster munition', it must satisfy the humanitarian effects *chapeau*, which will be presumed if the following criteria are cumulatively fulfilled: (i) each *munition* contains fewer than 10 explosive submunitions;

per (sensor-fused) weapon, in order to help 'greatly limit' the risks posed by unexploded submunitions (UXO) and to help avoid the incidence of indiscriminate attack.[292] The *CCM Commentary* explains that this works by restricting the number of explosive items that can distribute explosive force and fragmentation across a pre-defined area, *despite their capacity to* '*detect and engage a single target object*'.[293] This is significant, as it suggests the reasoning is two-fold: a) there is likely to be a limit to how many clearly detectable military objects there are in a relatively tight spatio-temporal boundary;[294] and, importantly for LAWS, b) the ATR of the submunitions can fail to accurately classify targets, and to distinguish military from civilian objects. The combination of these two will potentially result in an indiscriminate attack.

Similarly, it may be prudent to impose an upper limit on the number of specific engagements that a LAWS platform may undertake in a single 'out-of-the-loop' TLC mission. Namely, as the number and diversity of targets being pursued – and the extent of comingling – all increase, so too will the likelihood that an engagement will lead to distinction failure, for a given ATR and weapon system. Thus, it is arguable that the twin aims of the CCM of a) greatly limiting the *probability* of UXO and b) avoiding *indiscriminate* area effects (the latter partly emanating from the risk of ATR distinction failure) may combine to suggest a limit to the number of LAWS engagements per 'out-of-the-loop' mission. This also supports the argument that autonomous lethal targeting should be restricted to a single 'attack', which may comprise multiple (though limited) acts of violence;[295] hence, the utility of applying UELs.

On the exact UEL figure per 'out-of-the-loop' mission, there is again no bright-line rule that can be set in advance, as there is in Article 2(2)(c)(i), CCM. Instead, the exact UEL per deployment should depend on three cumulative sets of factors.

---

and, in turn, each explosive *submunition* (ii) weighs more than four kilograms; (iii) is designed to detect and engage a single target object; (iv) is equipped with an electronic self-destruction mechanism; and (v) is equipped with an electronic self-deactivation feature.

[292] Bonnie Docherty et al. 'Article 2: Definitions' in Gro Nystuen and Stuart Casey-Maslen (eds.), *The Convention on Cluster Munitions: A Commentary* (OUP, 2010), ¶¶ 2.121 and 2.129.

[293] Ibid., ¶ 2.124.

[294] There are major spatio-temporal differences between a sensor-fused weapon and a LAWS. Ultimately, however, deploying more force and target engagement capacity *in advance* and *without contemporaneous human oversight* arguably results in a greater likelihood of striking civilians or other protected persons and objects.

[295] See 4.3.

- The sophistication and reliability of the ATR and the control software. This may lead to upper engagement *guidelines* set by engineers, who will have regard to both the predetermined confidence thresholds and failure rates that were established during testing and evaluation. Guidelines may also take into account the platform's weapon carriage capacity and its range of damage potentials.
- The prevailing conditions in a given deployment, for example, the extent of clutter and comingling on the battlefield; the 'uniqueness' of the applicable target (categories); and the opportunities for multisensory phenomenologies, for detection-confirmation. This may lead to commanders and their battle staffs applying and adapting the guidelines set by engineers, to suit the operational environment and the mission and task at hand.
- The damage potential of the system on a given deployment, by way of the level and lethality of its armaments; whether these are anti-personnel or anti-material; the magazine depth; and the potential speed of engagements.[296] These may lead to commanders and their battle staffs further adapting engineers' guidelines.

Importantly, a LAWS will not necessarily have to return to base once the engagement limit is reached. Instead, UELs may be combined with an interim *review of performance* by way of a battle-damage assessment (BDA) and a system diagnostic check at the end of each cycle.[297] This would arguably provide a human buffer against distinction failure and indiscriminate attack, with weapons operators granting a 'permission to continue' if it appears that all engagements have been LOAC-compliant. Conversely, if problems are discovered mid-deployment, the operator should intervene in a way that is commensurate with the nature and extent of that problem. Options may include: lowering the UEL, hence increasing the frequency of human review; reverting the system to remotely-piloted mode; withdrawing the problematic LAWS from deployment; or withdrawing the entire fleet.[298]

---

[296] Paul Scharre, 'Autonomous Weapons and Operational Risk', *CNAS Ethical Autonomy Project* (February 2016), 18-19 <http://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf> accessed 22 September 2018.

[297] Recall from 4.4 that BDAs are a form of MHC. Recall also from 5.3.1.6 and 5.3.2.1 that the BDA is the final step in both Joint Targeting Cycles.

[298] This would be appropriate where poor performance is caused by a common component or feature of the system, whereby all systems exhibit the same failure mode at the same time.

In this sense, the human operator will act as a 'fail safe';[299] hence, UELs will arguably enhance the MHC afforded by tightly-constrained target parameters and spatio-temporal limits. Above all, a judiciously-set UEL will help to keep LAWS deployments within the 'individual attack' limitation.[300]

7.3.6.7 Training

Much of the above precautionary measures implicate the need for commanders and their battle staffs to undergo LAWS-specific LOAC training. As Corn points out, "training is the fundamental foundation upon which the soldier forms his judgment as to what is and is not permissible conduct during combat".[301] It should therefore be seen as both a prerequisite for the implementation of effective precautions, and as a precautionary measure in and of itself.

To illustrate the significance of LOAC training more broadly, *Directive 5100.77*[302] requires every member of the US Armed Forces to receive annual LOAC training, to enable them to fulfil their assigned functions in a legally compliant manner.[303] This is to maximise the chances of legally sound military operations and to minimise the risk of IHL violations; to the extent that LOAC training is now incorporated in virtually all classroom and field training exercises.[304] As Corn points out, such pervasive legal training reflects the fact that LOAC implementation demands more than mere knowledge of black-letter law.

> "[M]ilitary personnel at every level must be provided the opportunity to test their battlefield judgment and develop a genuine understanding of the relationship between the law and the execution of their mission through the crucible of realistic training."[305]

---

[299] See 4.2.1.

[300] This is more likely to be an issue in TLC; by definition, each TS is an individual attack, subject to reasonable spatio-temporal limits.

[301] Corn (n 231), 445.

[302] US Department of Defense, *Directive No. 5100.77: DoD Law of War Program* (9 December 1998) (hereafter, *Directive 5100.77*) <https://biotech.law.lsu.edu/blaw/dodd/corres/pdf2/d510077p.pdf> accessed 22 September 2018.

[303] Ibid., §§ 4 (Policy) and 5 (Responsibilities).

[304] Corn (n 231), 446.

[305] Ibid., 446-47.

Moreover, in the case of LAWS, training is regarded as a MHC touchpoint, which ensures the input of *effective* controlled processing during the targeting cycle.[306] Accordingly, for commanders and their battle staffs to maximise the front-loading of critical tasks; to activate online learning in a safe manner; to set prudent spatio-temporal and target parameters; and to implement judicious UELs, a strong element of applied LOAC training will arguably be needed *before* autonomous weaponry is fielded and ready for deployment. Given the highly contextual nature of these precautionary measures and the absence of any bright-lines, targeting personnel will need, first, to appreciate the capacities and limitations of the LAWS that they are to deploy; and, second, to develop their judgment in a range of realistic training scenarios, where inadvertent errors become learning opportunities rather than fatalities. Only once these skills are developed and refined through the precautionary step of applied LOAC training can LAWS be deployed with the confidence that the precautionary measures developed during the targeting cycle will be fit for purpose.

7.3.6.8 Staffing

While a lay understanding of capacities and limitations and how to prudently apply precautionary measures is usually sufficient for final decision-makers, the *overall* targeting process will arguably need more than this. To be sure, the more complex and stochastic the weapon system, the greater the need for roboticists and software engineers to be integrated into the battle staffs.[307] In this connexion, the highly technical nature of LAWS and their likely brittleness will arguably give rise to information asymmetries and safety risks, which can only be resolved by integrating appropriate technical personnel.[308]

Currently, Article 82, AP I, requires States to ensure that commanders have access to legal advisers before making decisions, but there is no specific requirement for technical personnel to give advice on performance and effects.[309] Without doubt, the integration of such personnel into the battle staffs would be a broad precautionary measure that is outside the scope of the narrow LOAC rules. Above all, it would

---

[306] See 4.4 (Table 4.1), 4.5.2 (commander understanding of technology) and 4.5.3 (operator training).

[307] Tony Gillespie, 'New Technologies and Design for the Laws of Armed Conflict' (2015) 160 The RUSI Journal 50.

[308] Alan Backstrom and Ian Henderson, 'New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews' (2012) 94 International Review of the Red Cross 483.

[309] Gillespie (n 307), 50.

provide commanders with access to indispensable expertise, in the face of strong information asymmetries.

On a practical note, recall from Chapter 5 that US doctrine on Joint Operations is organised around 'Joint Functions'. With the recent addition of *Information*, partly to cater for the increasing complexity of software in weapon systems,[310] there are now seven Joint Functions.[311] Within these, roboticists and software engineers could arguably fit into *Fires* and *Information*, to provide the technical input that lay commanders would take into account when deciding on deployments, and when formulating LAWS-specific precautionary measures.

7.3.7 How to 'Elevate' Precautions to a Full LOAC Principle

While none of the eight precautionary measures suggested above are specifically mandated by law, all of them may conceivably flow from a broader precautionary principle.[312] To this end, we may query *how* such a recasting of the precautionary norms, from a set of narrow rules to a generally applicable principle, may be effectuated. As noted in 7.3.5.2, this should not require a change in the law as such, as the 'constant care' obligation already provides a legal basis for a precautionary principle.[313] Instead, the challenge of 'elevating' precautions should be regarded as an effort to enhance compliance with LOAC.[314]

Certainly, in a US/NATO context this may be expected to be relatively straightforward because of Joint Targeting Doctrine, which already infuses the targeting process with a strong precautionary character. To increase the likelihood that this will be optimised

---

[310] More precisely, it is an acknowledgment that virtually all of the work of the Armed Forces is now driven by information networks, especially in space and cyberspace operations.

[311] Joint Chiefs of Staff, *Joint Publication 1: Doctrine for the Armed Forces of the United States* (25 March 2013, incorporating Change 1, 12 July 2017), I-17–I-19 (detailing the Joint Functions: *C2*; *Intelligence*; *Fires*; *Movement and Manoeuvre*; *Protection*; *Sustainment*; and *Information*).

[312] Though some of these measures may also flow from efforts to comply with the precautionary rules. For example, narrowing the target parameters, as described in 7.3.6.5, may be regarded as essential for complying with the target verification rule in Article 57(2)(a)(i), AP I.

[313] Moreover, with the use of the auxiliary verb "shall" in Article 57(1), AP I, the binding status of this obligation is beyond question. The same reasoning applies to the ICRC's restatement of customary law, with CIHL, Rule 15, stating that "constant care *must* be taken…" (emphasis added).

[314] In the case of some non-NATO States, there is arguably also a need for advocacy in relation to their interpretation of "constant care". See Jean-Françoise Quéguiner, 'Precautions Under the Law Governing the Conducting of Hostilities' (2006) 88 International Review of the Red Cross 793, 796 (noting the tendency of some States and commentators to see the constant care obligation as "merely inspirational", because it forms "a sort of preamble to Article 57", and because of its "very general wording").

for LAWS deployments, a three-pronged and iterative approach is proposed. First, a pervasively precautionary doctrine on autonomous attack can be developed and documented in Joint Doctrine Publications. These can outline some LAWS-specific *measures* (such as those examined in 7.3.6.1-7.3.6.8) and they can detail the *process* of developing further specific measures for a given deployment.[315] This will provide commanders, their legal advisers and their battle staffs with a blueprint for how to put into practice a genuine precautionary principle for LAWS deployments.[316]

Second, all commanders, legal advisers and battle staffs would need to undergo LAWS-specific LOAC training, to ensure the finer details of the new doctrine is fully absorbed and correctly translated into targeting practice. This is already required of US forces by *Directive 5100.77* and the reasoning for it is outlined in 7.3.6.7, above. Much of what is discussed there also applies here, with the addition that applied LOAC training will instil in all personnel an understanding that 'precautions' is a broad principle, and that doctrine and process are a deliberate means to realise this.[317]

Third, there will arguably be a need to identify and share best practices between armed forces.[318] Across NATO, and even within NATO States, military structures tend to differ,[319] and this is likely to negate a one-size-fits-all approach.[320] Yet, such idiosyncrasies do not justify an erosion of the process of LOAC implementation, nor any disintegration of the precautionary principle. This arguably calls for best practices

---

[315] In preparing these, an extensive Joint Doctrine development process will need to be undertaken for entirely new doctrine publications. Shorter and/or expedited processes will suffice for LAWS-specific updates to other (general) doctrine publications. See North Atlantic Treaty Organisation, *AAP-47: Allied Joint Doctrine Development* (Edition B Version 1, NATO Standardisation Office, June 2016); Joint Staff, *Joint Doctrine Development Process*, CJCSM 5120.01A (Joint Staff, 29 December 2014).

[316] As alluded to in the previous footnote, such documentation would ideally consist of a dedicated Joint Doctrine Publication on LAWS deployments; as well as updates and amendments to existing publications, such as those on Joint Targeting that were cited in Chapter 5.

[317] As noted in 7.3.6.7, training is therefore an essential means for elevating precautions to a full LOAC principle, and it is a precautionary measure in itself.

[318] Jakob Kellenberger, *Strengthening Legal Protection for Victims of Armed Conflicts: States' Consultations and Way Forward* (ICRC, 12 May 2011), 2 (noting that ways to strengthen LOAC compliance include "the elaboration of soft law instruments, *the identification of best practices* and the facilitation of expert processes aimed at clarifying existing rules") (emphasis added) <https://www.icrc.org/en/doc/assets/files/red-cross-crescent-movement/31st-international-conference/icrc-president-statement-2010-05-12.pdf> accessed 26 March 2019.

[319] Corn (n 231), 429.

[320] Indeed, as the *AP I Commentary* points out, at ¶ 3344, this is one reason why the Article 82 obligation to provide commanders with legal advice is qualified by allowing States to decide the level of command at which to integrate legal advisers.

to be identified and "marketed" to less developed armed forces, to promote a more uniform commitment to implementing the precautionary principle.[321]

Finally, once training is established, LAWS are being actively deployed, and best practices have been identified and disseminated, it is likely that some actual or desired changes to the original doctrine will be discovered. If so, Joint Doctrine on LAWS should be updated and its associated Publications should be amended to reflect this, thereby continuing to 'elevate' precautions in an iterative manner.

The above approach is arguably feasible for its focus on a limited number of relatively like-minded (NATO) States, which are already predisposed to treating precautions as a principle. However, as LAWS are likely to be fielded by other (non-NATO) States too, it is preferable to share best practices more widely and to facilitate expert processes aimed at clarifying the application of the constant care obligation.[322] Moreover, as some non-NATO States may not even regard precautions as being a broader principle,[323] it will also be necessary to clarify and advocate the correct legal interpretation, to ensure those States will be receptive to any sharing of best practices.

7.4 Conclusion

Proportionality is clearly the most difficult of LOAC norms for a LAWS to satisfy through automatic processing alone. Yet, we must not exaggerate the difficulties, as many aspects of the principle that require controlled processing can be delegated to the commander during the targeting process, or to a weapons operator acting through a dial-in capability. If neither option is available, deployments will simply have to be restricted. More useful for LAWS are the precautions in attack rules, both as tasks to be front-loaded for controlled processing and/or as machine actions that can be resolved through automatic processing in the field. However, the listed precautions in AP I are limited in scope and are not necessarily apt for LAWS operations. An arguably better approach is to harness precautions as a full LOAC principle, to maximise the range of LAWS-specific precautionary measures that will be fit for

[321] Corn (n 231), 429; also arguing, at 452, that relevant best practices should relate to a) integrating legal advice into the targeting process, and b) ensuring the process itself maximises LOAC compliance.

[322] Kellenberger (n 318), 2. See also 8.2, on the merits of producing a LOAC Manual on LAWS.

[323] See Quéguiner's comments on how some States see the constant care obligation, at n 314.

purpose in the circumstances of each deployment.[324] This will increase the likelihood of securing compliance with the principles of distinction and proportionality. Accordingly, before LAWS are fielded it is crucial that US/NATO forces begin developing and applying a broader precautionary principle for autonomous attack.

[324] In addition to the eight broader precautionary measures illustrated above, the use of a kill switch with a feedback loop would also be an essential precautionary feature of a LAWS deployment. This was identified as an aspect of MHC in 4.4 and Table 4.1 ('use and abort'), 4.5.2, 4.5.4 (timely intervention) and 4.5.5 (feedback loop and abort). On mandating kill switches via legal transplants from the landmines regime, see Homayounnejad (n 270), 152-54.

Chapter 8
Conclusion

With numerous States actively developing lethal autonomous weapon systems (LAWS) and the world's largest defence bureaucracy (US Department of Defense) integrating autonomy into its latest *National Defense Strategy*,[1] it is arguably likely that offensive autonomous capabilities will soon appear on the battlefield. Add to that the rapid pace of technological change,[2] the return to Great Power Competition,[3] and the potential offence-defence dynamic,[4] and the prospect that some States will field and deploy LAWS in the near-term looks almost inevitable. The consequent substituting of algorithmic targeting for narrow-loop human targeting sounds bleak, especially when it concerns matters of life and death. However, as has been demonstrated in this thesis, there are means and ways to make autonomous attack both lawful and relatively humane, and of sufficient military utility as to incentivise armed forces to act lawfully. Yet, this is by no means a guaranteed outcome: it will require a carefully considered application of the laws of armed conflict (LOAC) in the context of an elaborate targeting process; fully informed by the human-machine cognitive differences, and firmly resisting any temptation to anthropomorphise the systems.

8.1 Linking Back to the Research Question

With this in mind, recall the research question laid down in 1.4:

> To what extent can US/NATO forces apply the existing LOAC targeting rules, to ensure the lawful deployment and use of near-term LAWS?

This arguably implies two sub-questions:

(1) To what extent can the existing LOAC targeting rules be applied to ensure the lawful deployment and use of LAWS?

(2) To what extent are US/NATO forces *in particular* able to undertake this challenge?

[1] See 1.1

[2] See 1.2.1.

[3] See 1.2.2.

[4] See 1.2.3.

In relation to (1), this thesis has demonstrated that the lawful deployment of LAWS will be possible if commanders can ensure that a) appropriate systems are deployed, b) in a suitable operational environment, c) to undertake machine-feasible tasks, along with d) appropriate precautionary measures to sufficiently mitigate civilian risk. Deployment A in Figure 8.1, below, shows all these elements equally present, collectively meeting the 'threshold of lawfulness' at a notional 100. However, where any one of these elements is lacking, the other(s) should be adjusted in some way to compensate. For example, where autonomous attack is planned in a relatively complex operational environment, commanders may still meet the threshold of lawfulness by deploying a more sophisticated system, for simpler (more machine-feasible) targeting, and with relatively stronger precautions in attack. This is illustrated in Deployment B.

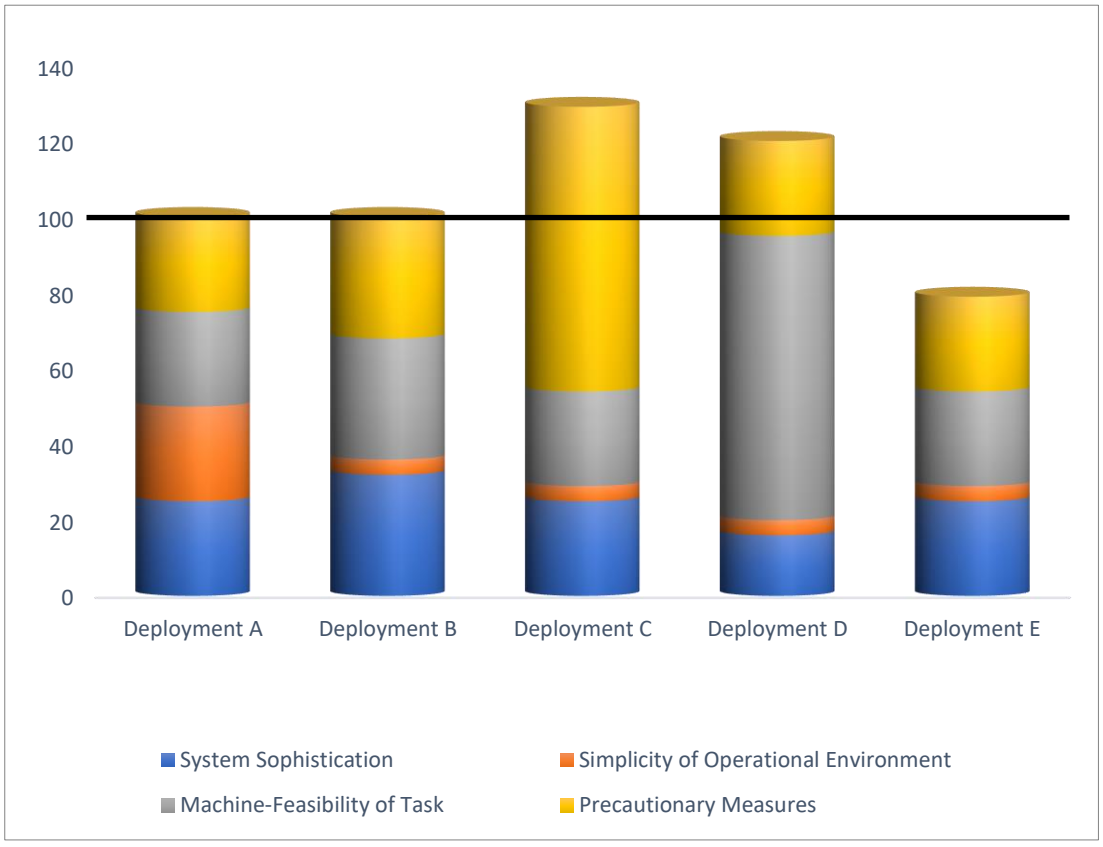

Figure 8.1: Alternative deployment scenarios and the 'threshold of lawfulness'.

Of course, there are no bright-lines and the precise balance of these four factors will be a matter for commanders to resolve in any given scenario, utilising their training, their common sense, their legal and battle staffs, and always acting in good faith. Accordingly, in Deployment C it may be more appropriate to keep the system and task constant, but to significantly boost the precautionary measures, to *over*compensate and

ensure the deployment stays on the right side of the margin of error. Meanwhile, Deployment D sees the commander opting for simpler systems restricted to the simplest and most machine-feasible (anti-material) targeting, again overcompensating for the complex operational environment. Ultimately, context is king, and it will be for the commander to decide how a LAWS deployment will meet (or exceed) the threshold of lawfulness *in the specific mission at hand*. Conversely, where the threshold is not met due to shortfalls that cannot be offset, then a LAWS cannot be lawfully deployed; as in Deployment E, which will have to go ahead with manned or remotely-piloted systems, or be cancelled altogether.

In relation to the second sub-question, it was seen in Chapter 5 that US/NATO forces operating under Joint Targeting Doctrine already have an effective and robust infrastructure in place to meet the 'threshold of lawfulness' challenge. Moreover, Chapters 6 and 7 demonstrated that this can indeed interact with targeting law in such a way as to ensure legally compliant LAWS deployments. Certainly, some adjustments to the targeting process will be needed, such as elevating precautions to a full LOAC principle, and integrating more technical personnel into the battle staffs. Furthermore, some current developments, such as *Project Maven* and any other form of wider loop autonomy, will need to be closely monitored to assess whether these do actually enhance human control during target development. Yet, these are not insurmountable goals; the existing Joint Targeting process arguably provides US/NATO forces with an effective starting point and a means to achieve them.

8.2 Beyond the US/NATO Context: The Value of Standardisation

Outside the US/NATO context, the above LAWS-specific approach may not always be obvious; certainly not to every commander in every potential situation in armed conflict. The result may be that some fail to implement it, particularly those from less advanced militaries that have a more limited provision of legal and technical staffs; or indeed inexperienced and poorly supported commanders coming under extraordinary pressures in armed conflict. Accordingly, as noted in 1.1, there have also been calls for specific regulation of LAWS, either through a binding treaty or a non-binding LOAC Manual, to at least set out a common understanding of how such weapon systems can be deployed and used lawfully.

We need not look too far for real-world conflicts that support this latter view. For example, consider the uses of white phosphorous, flechettes, and (allegedly) DIME[5] weapons during the Gaza Conflict of January 2009. Such weapons are neither illegal *per se* nor are they specifically regulated,[6] but their deployment and use are still subject to the generally applicable rules and principles of LOAC. Yet, during the Gaza Conflict these weapons may have been used unlawfully; for example, by being deployed in areas containing a concentration of civilians.[7] Arguably, this occurred partly *because* of the lack of specific regulation of these weapons, which left the application of the law in the hands of individual commanders, whose legal assessments were in stark contrast to those of human rights groups.[8]

Along similar lines, there may be a case to specifically regulate, or provide soft-law guidance on, the deployment and use of LAWS. Indeed, given the novelty of lethal autonomy – especially when combined with long loitering capabilities – rules specifically crafted for such weapons will arguably be a welcome step in the development and application of LOAC. However, the utility of this approach goes beyond merely helping to standardise the practice of non-NATO forces. It may also support efforts *within* NATO to 'elevate' precautions to the status of a full LOAC principle in a LAWS context. Namely, a treaty or LOAC Manual may act as an effective precursor to the kinds of implementation steps explained in 7.3.7.

Thus, with a LAWS regulation treaty negotiated by the States, or a LOAC Manual formulated by a group of experts, we should move closer to a common understanding of the deployment and use of LAWS in full compliance with the rules that regulate the conduct of hostilities; both in a NATO context, and in a broader global context. This now raises the question as to which of the two instruments – treaty or Manual – might be preferred, and for what reasons.

---

[5] Dense Inert Metal Explosive.

[6] Although there is specific regulation of munitions containing white phosphorous if they are "primarily designed to set fire to objects or…cause burn injury to persons": Article 1(1), Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons (adopted 10 October 1980, entered into force 2 December 1983) 1342 UNTS 171.

[7] See UN Human Rights Council, *Report of the United Nations Fact-Finding Mission on the Gaza Conflict* (25 September 2009) UN Doc. A/HRC/12/48, especially 191-98 <http://www2.ohchr.org/english/bodies/hrcouncil/docs/12session/A-HRC-12-48.pdf> accessed 13 October 2018.

[8] See, for example, Harriet Sherwood, 'Israel Using Flechette Shells in Gaza', *The Guardian* (20 July 2014) <https://www.theguardian.com/world/2014/jul/20/israel-using-flechette-shells-in-gaza> accessed 13 October 2018.

8.2.1 The Normative Status and Value of a Treaty *versus* a LOAC Manual

Recall from 1.3.1 that treaties and LOAC Manuals each have a different normative status under international law. Treaties – be they standalone instruments or a new Protocol under the Convention on Certain Conventional Weapons[9] (CCW) – are a recognised source of international law; thus, they enjoy binding status.[10] By contrast, LOAC Manuals are a "subsidiary means for the determination of rules of law";[11] hence, they offer valuable interpretive guidance, but they are *non*-binding. On this basis alone, a treaty may be considered to be more valuable for LOAC compliance.

On the other hand, treaties are often less comprehensive than LOAC Manuals, as States rarely wish to bind themselves to detailed and rigid legal provisions, especially in evolving fields. This will arguably reduce the utility of a LAWS regulation treaty, as it may only be possible to conclude such an agreement on such a broad category of systems and deployment contexts by using vague or heavily caveated provisions, along the lines of Protocol V to the CCW.[12] Even where we do see precision and detail in Protocol V, this only relates to its non-binding Technical Annex, which contains 'voluntary best practice'.[13] Moreover, this instrument and its Annex stretch to no more than 14 pages, whereas LOAC Manuals tend to be more comprehensive, often with hundreds of pages of details and guidelines. For example, the *AMW Manual*[14] consists of 56 pages, while the *Tallinn Manual 2.0*[15] contains a staggering 598 pages. Free from the political pressure of States trying to avoid rigid legal positions, these Manuals can offer relatively greater insight and guidance on applying LOAC in a given domain, leaving it to the commander to apply the rules in the context of a specific deployment.

---

[9] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects (adopted 10 October 1980, entered into force 2 December 1983, amended on 21 December 2001) 1342 UNTS 137.

[10] Article 38(1)(a), Statute of the International Court of Justice (adopted 26 June 1945, entered into force 24 October 1945) 145 BFSP 832 (hereafter, ICJ Statute).

[11] Article 38(1)(d), ICJ Statute.

[12] Protocol on Explosive Remnants of War (Protocol V to the 1980 CCW) (adopted 28 November 2003, entered into force 12 November 2006) 2399 UNTS 100 (frequently caveating its provisions with "where feasible", "to the maximum extent possible", and "as far as practicable").

[13] Moreover, under Article 9, Protocol V, States are only "encouraged" to take "generic preventive measures aimed at minimising the occurrence of explosive remnants of war", meaning that even consultation of the Technical Annex is not mandatory.

[14] Program on Humanitarian Policy & Conflict Research at Harvard University, *HPCR Manual on International Law Applicable to Air and Missile Warfare* (Harvard College, 2009).

[15] Michael N. Schmitt (ed.), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd ed., CUP, 2017).

An alternative scenario might involve a hard core of States and non-governmental organisations (NGOs) pressing ahead with a more detailed set of treaty proposals, with the intention that other States will eventually follow suit. This would provide a combination of detail and binding status. However, with the current CCW membership strongly in disarray over how to move forward with LAWS, and some States remaining rather unresponsive to NGO pressures,[16] even this may be unlikely to work. Instead, such an approach risks turning LAWS into a "Balkanized sector of weapons law",[17] with significant LAWS-deploying States likely to opt out. Indeed, this is how the Ottawa Process ended: despite leading to the comprehensive Mine Ban Treaty,[18] States like the US, Russia, China, and South Korea all refused to sign and/or to take part, even while they mostly comply on a voluntary basis.[19] Of course, a LAWS regulation treaty would not go so far as to ban autonomous weaponry, but would instead clarify its terms of deployment and use. Yet, the countervailing context is equally important: anti-personnel mines had already lost some of their military utility by the time the Ottawa Process was underway, thus they were relatively susceptible to a treaty ban. By contrast, many States proclaim the potential military utility of LAWS, so it is likely that even more of them will refuse to sign up to a detailed instrument that ties their hands in advance; hence, the reference to the 'Balkanization' of weapons law.

Once again, we arrive at the conclusion that an expert LOAC Manual on LAWS is the approach that is most likely to succeed and be effective, because of its combination of flexibility, detail, and specificity. In turn, the Manual is relatively more likely to see widespread adoption and effective compliance by LAWS-deploying forces.

---

[16] See, for example, Denise Garcia, 'Governing Lethal Autonomous Weapon Systems', *Ethics & International Affairs* (13 December 2017) <https://www.ethicsandinternationalaffairs.org/2017/governing-lethal-autonomous-weapon-systems/> accessed 10 June 2018 (explaining that State parties are split along three major lines: those wanting a ban or moratoria; those opposing a ban, or even any specific regulation; and those advocating a politically binding agreement based on meaningful human control concepts, but no legal solution).

[17] Sean Watts, 'Autonomous Weapons: Regulation Tolerant or Regulation Resistant?' (2016) 30 Temple International & Comparative Law Journal 177, 187.

[18] Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction (adopted 18 September 1997, entered into force 1 March 1999) 2056 UNTS 241.

[19] A similar situation exists with the Oslo Process that led to the Convention on Cluster Munitions (adopted 30 May 2008, entered into force 1 August 2010) 2688 UNTS 190. States like the US, Russia, China, and South Korea either refused to sign up or even to participate in this NGO-led process. In part, they objected to the stringent technical design criteria – now codified in Article 2(2)(c) – which these States saw as depriving them of legitimate military utility to be gained from using 'cluster munitions'.

The overall value of a LOAC Manual on LAWS can therefore be listed as follows:

- It will provide an authoritative source of *expert interpretive guidance*, specifically on applying international law to a new generation of weapon systems, and with an appropriately interdisciplinary focus.
- It will *fill gaps in treaty law* relatively quickly, without the political or procedural complexity of the treaty-making process.
- Due to its non-binding status, the rules, principles and guidance of a LOAC Manual can go into *considerable detail*, offering commanders and their legal advisers an invaluable source of expert insight.
- Also due to its non-binding status, a LOAC Manual is *more likely to be adopted by the world's militaries*, which will apply its rules, principles and guidance with appropriate flexibility.
- Collectively, the above will *promote a more standardised approach* to the deployment and use of LAWS for legal compliance, both within and outside NATO.

Over time, such standardisation may also foster the development of a new customary international law on autonomous lethal targeting; even more so if States copy the LOAC Manual's rules, principles and guidance into their own national military manuals.

8.2.2 The Potential Contours of a LOAC Manual on LAWS

With the above in mind, we can now address the substantive headings that may appear in a LOAC Manual on LAWS. These will all benefit from a strongly interdisciplinary input from recognised experts on robotics, software engineering, cognitive sciences, weapons design and human-machine interaction, and international law.

The Manual can be divided into eight broad chapters. First, a **Definitions** section can set forth key LAWS-specific definitions, such as 'autonomy' and 'online learning' *versus* 'offline learning', amongst others. It can also set out key LOAC definitions, such as 'military objective', 'collateral damage', and 'military advantage'. Crucially, this section can also include a detailed account of the 'meaningful human control' (MHC) standard and its potential elements, for reference in subsequent sections.

Second, a **General Framework** section can affirm that LAWS are governed by existing LOAC (and other international laws); that they are not inherently unlawful; yet the fundamental principle remains that the right of the Parties to armed conflict to choose their methods or means of warfare is not unlimited; hence, LAWS are subject to the customary law requirement of legal review.[20] This section can also explicitly affirm that the Manual is a restatement of existing treaty and customary law, and that it operates without prejudice to the existing treaty obligations of States.

Third, a **Weapons Law** section can restate the rules that prohibit weapons whose very nature or design leads to unlawful results;[21] and it may provide guidance, or restate a prohibition, on the use of online learning.[22] This section can also include rules and policy guidelines for LAWS development; for example, it may address whether all systems should be equipped with a 'kill switch' or a built-in self-neutralisation mechanism. It may also address minimum sensory requirements and processing capabilities for LAWS that could be required to make reasonably foreseeable distinction-based determinations.[23] Moreover, the section can reaffirm the customary law requirement of legal review and it can suggest some broad principles in relation to this, such as the need for interdisciplinary review panels.

Fourth, a ***Jus ad Bellum*** section can address situations where the use of LAWS may affect the initial recourse to force. This would take the Manual beyond LOAC issues; however, as with the *Tallinn Manual 2.0*, such a broadening of scope is justified and appropriate where the weapon system may genuinely pose risks to the *ad bellum* framework, or otherwise obscure its application.[24] In a LAWS context, the canonical example of an *ad bellum* complication is the risk of 'flash war'.[25] This may occur where opposing fleets of LAWS, deployed during peacetime, meet very closely in time and space and observe each other on high alert, while looking for any sign of an

---

[20] Kenneth Anderson, Daniel Reisner and Matthew C. Waxman, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems' (2014) 90 International Law Studies 386, 407.

[21] For example, the rules prohibiting superfluous injury and unnecessary suffering, environmental damage, inherently indiscriminate operation, and uncontrollable effects; under Articles 35(2)-(3) and 51(b)-(c), AP I, respectively. See 6.3.

[22] See 2.5.1.4.

[23] Anderson, Reisner and Waxman (n 20), 407.

[24] See *Tallinn Manual 2.0*, Part III: International Peace and Security and Cyber Activities.

[25] See Paul Scharre, 'Autonomous Weapons and Operational Risk', *CNAS Ethical Autonomy Project* (February 2016) <http://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf> accessed 22 September 2018.

impending attack. In such a scenario, systems from State A may well misinterpret the (potentially innocuous) manoeuvres of systems deployed by State B, triggering a swift and sudden military offensive, followed by a counter-offensive and rapid escalation into high-intensity conflict.[26] If and when this occurs, it may leave great uncertainty as to whether the initial shot was an unlawful use of force or a violation of territorial integrity;[27] or a legitimate act of (anticipatory) self-defence.[28] A Manual could address this by recommending safeguards to prevent such algorithmic calamities. For example, it may consider circumstances that call for relatively tighter attack parameters; a 'shoot second' policy; and, potentially, the sharing of certain non-classified algorithmic features, to bring adversarial systems closer to 'peacetime compatibility'.

Fifth, a **Targeting Law** section can address many of the same issues that were raised in Chapters 6 and 7 of this thesis: how to deploy LAWS in compliance with the rules and principles of distinction, proportionality, and precautions in attack. Much of the analysis put forward in this thesis may therefore provide an input; however, as LOAC Manuals are meant to express the *consensus* of experts, it is likely that the analysis in Chapters 6 and 7 would be subject to further input and amendment, in order to reach such consensus.

For this hypothetical fifth section, a LOAC Manual could usefully incorporate the following features:[29]

- Clarifying that decisions regarding distinction and proportionality are made by commanders and weapons operators *not* the LAWS; and that such decisions are generally moved forward in time, to the point of deployment or earlier (unless there is a 'dial-in' capability).[30]

---

[26] Jürgen Altmann and Frank Sauer 'Autonomous Weapon Systems and Strategic Stability' (2017) 59 Survival 117, 128.

[27] Article 2(4), Charter of the United Nations (adopted 26 June 1945, entered into force 24 October 1945) 1 UNTS XVI (hereafter, UN Charter).

[28] Article 51, UN Charter; *The Caroline Case* (1841) 29 BFSP 1137. With digital data-recording, it will be relatively easy to determine which side took the first shot, but more difficult to determine whether that first shot was reasonable in the circumstances; hence, whether it was an unlawful use of force or a legitimate act of (anticipatory) self-defence.

[29] Some of these are also put forward in Anderson, Reisner and Waxman (n 20), 407; and Steven Groves, 'A Manual Adapting the Law of Armed Conflict to Lethal Autonomous Weapons Systems', *Margaret Thatcher Center for Freedom, Special Report No. 183*, (The Heritage Foundation, 7 April 2016), 4-6 <http://thf-reports.s3.amazonaws.com/2016/SR183.pdf> accessed 4 October 2018.

[30] This will help to avoid anthropomorphisms, in order to retain MHC; and it will avoid unhelpfully dystopian narratives of 'killer robots' making their own 'decisions' and running amok, which often distracts from a lucid analysis of legal compliance.

- Providing an interpretive application of LOAC to LAWS, explaining a) what information commanders must have and b) what questions they must generally ask before and during deployment.
- Setting out realistic deployment scenarios, from the benign to the highly complex, and considering relevant elements of the MHC standard *needed to secure compliance with LOAC norms*.[31] This can usefully be placed within the framework of the 'threshold of lawfulness' challenge laid out in 8.1.
- Restating the importance of the 'constant care' obligation and its implication for a LOAC precautionary principle; and applying this in the LAWS context.[32]
- Considering situations in armed conflict where human rights norms will assume greater significance, hence where a stronger element of contemporaneous human judgment and control will be necessary, even to the point where a LAWS will have to be shut down or remotely-piloted.

Importantly, the above should also consider how the rules and their application to LAWS might vary between international and non-international armed conflict.

Sixth, an **Accountability and Responsibility** section can restate and apply the various rules on (individual and command) responsibility in international criminal law; this may extend beyond combatants and commanders, to include software programmers, designers, and manufacturers. Separately, the section may also link to the previous one, to clarify the levels of technical knowledge (of system capacities and limitations) needed by a weapons operator or commander, to establish them as an accountable person in LOAC; again, this may extend to non-traditional entities, such as legal review panels and procurement teams. Finally, this section can restate and apply the relatively non-controversial rules on State responsibility, for when any use of LAWS by a State's armed forces leads to unlawful damage, death or injury; even where such harm was unpredictable or due to a malfunction.

Seventh, a **Law of Neutrality** section can restate and apply to LAWS the rules that preserve the inviolability of the territory of neutral States; and the rights and obligations of those neutral States in relation to LAWS-deploying Belligerents.

---

[31] As opposed to regarding human involvement as an independent legal requirement. See 4.5.6.

[32] The aim would be to instil a pervasively precautionary approach that fosters the creation of apt, LAWS-specific precautionary measures in every deployment.

Finally, a **Law of Occupation** section can restate and apply the rules relating to respect for protected persons and property in occupied territory; public order and safety; and the security of the occupying authority, amongst others. Importantly, as there is a stronger role for human rights norms during occupation, this section may restate a requirement for relatively tighter attack parameters, non-lethal operation, and/or contemporaneous human judgment and control in some deployments.

The above only represents the broad contours; nonetheless, it does suggest that a relatively comprehensive LAWS Manual is, on the face of it, viable.

8.2.3 Potential Challenges to Developing a LOAC Manual on LAWS

Yet, on closer examination, such an undertaking will not be easy: unlike previous LOAC Manuals, such as *AMW* and *Tallinn*, a LAWS Manual would be developed and published in a hostile political environment.[33] As discussed in 1.1, there is a growing campaign led by NGOs and, increasingly, some States to pre-emptively ban LAWS. This will raise a significant political challenge to the broad acceptance of any permissive norms restated in a LOAC Manual. Namely, as the momentum amongst these stakeholders is to ban and not normalise LAWS, there is a clearly suboptimal environment for developing a widely-accepted LAWS Manual.[34] Not helping the situation is the fact that the International Committee of the Red Cross (ICRC) has remained rather reticent on the LAWS issue, and has even voiced some concerns.[35] This may indicate its reluctance to support a LAWS Manual, which would be in contrast to previous projects, such as the *San Remo Manual* on naval warfare.[36] There, the ICRC was supportive throughout, and this gave the *San Remo Manual* a strong air of legitimacy, largely because of the organisation's status as an impartial, neutral and independent body with an exclusively humanitarian mission. Accordingly, the ICRC's apparent lack of enthusiasm for the potential benefits of machine autonomy may further undermine the prospect of gaining broader acceptance for a LAWS Manual.

---

[33] Groves (n 29), 8 (noting that other LOAC Manuals benefited from an absence of NGO resistance).

[34] Ibid.

[35] ICRC, 'Autonomous Weapon Systems – Q&A', *ICRC Article* (12 November 2014) <https://www.icrc.org/en/document/autonomous-weapon-systems-challenge-human-control-over-use-force> accessed 10 June 2018 (stating that while it "has not joined these calls [for a ban] for now", the ICRC remains concerned about "allow[ing] machines to make life-and-death decisions", as this "would reflect a paradigm shift…in the conduct of hostilities". Thus, it cautions against the use of such weapons, unless LOAC compliance can be "guaranteed", which is a very high – arguably unattainable – standard).

[36] Louise Doswald-Beck (ed.), *San Remo Manual on International Law Applicable to Armed Conflicts at Sea* (CUP, 1995), 5 (crediting the ICRC for its supportiveness, in line with its humanitarian mandate).

That said, the ICRC has not officially joined the call for a ban.[37] Instead, the organisation has become more active in researching and convening meetings of technical experts on human-machine interaction.[38] This has resulted in some extensive contributions to the debate on human control and its determinants,[39] which *may* indicate a cautious willingness to be involved in developing the normative framework for LAWS deployments.

8.2.4 A Possible Way Forward

The issue of LAWS is clearly divisive, and the rapidly evolving nature of its technology brings numerous future uncertainties. However, if we accept the inevitability of LAWS; that these machines can potentially be used lawfully; and that appropriately planned deployments may result in fewer civilian casualties, then the best response to division and uncertainty is arguably a non-binding LOAC Manual. Building on the success of other Manuals – *San Remo*, *AMW*, and *Tallinn*, to name just a few – an organisation like NATO could convene a group of experts drawn from the armed forces, legal academia, robotics, software engineering, cognitive sciences, and weapons design. Ideally, this group would have equal representation from all States, but it is likely that most experts in robotics, weapons design, and international law will be drawn from a small number of developed nations.[40] Moreover, some States that have called for a ban are likely to be reluctant to send any State-appointed experts, or even to recognise the project, while NGOs are likely to actively campaign against it. For this reason, it is even more important that the convenor enlists the ICRC, to bring a greater sense of legitimacy to the LAWS Manual process. Despite its apparent reticence, the ICRC may be cautiously willing to be involved, if only because it recognises the inevitability of LAWS and may see the opportunity to infuse the process with a stronger humanitarian approach. This should arguably be welcomed, so long as it maintains and does not displace the delicate balance between military necessity and humanitarian concerns. Above all, the involvement of the ICRC may help to gain broader acceptance for a Manual that will very likely be swimming against a tide of NGO resistance.

---

[37] ICRC (n 35).

[38] ICRC, 'Statement of the International Committee of the Red Cross Under Agenda Item 6(b)', *ICRC Statement Delivered to the 2018 GGE Meeting on LAWS* (27-31 August 2018) <https://www.unog.ch/80256EDD006B8954/(httpAssets)/151EF67AD8224E14C125830600531382/$file/2018_GGE+LAWS+2_6b_ICRC.pdf> accessed 30 September 2018.

[39] See 4.5.4, and the references contained therein.

[40] Groves (n 29), 11. To this should be added Russia and China, because of their advanced militaries.

Bibliography

Books and Book Chapters

Anderson K and Waxman MC, 'Debating Autonomous Weapon Systems, their Ethics, and their Regulation under International Law' in Brownsword R, Scotford E and Yeung K (eds), *The Oxford Handbook of Law, Regulation and Technology* (OUP, 2017)

Arkin RC, *Governing Lethal Behaviour in Autonomous Robotics* (Chapman & Hall/CRC, 2009)

Arquilla J and Ronfeld D, *Swarming and the Future of Conflict* (RAND, 2000)

Badar ME, *The Concept of Mens Rea in International Criminal Law: The Case for a Unified Approach* (Hart Publishing, 2015)

Belcher ML Jr and Scheer JA, 'Radar System Implementation' in Whitaker JC (ed), *The Electronics Handbook* (2nd ed, CRC, 2005)

Best G, *Humanity in Warfare: The Modern History of the International Law of Armed Conflicts* (Methuen, 1983)

Bhosal B, 'Curvelet Interaction with Artificial Neural Networks' in Shanmuganathan S and Samarasinghe S (eds), *Artificial Neural Network Modelling* (Springer International, 2016)

Bianchi A, *International Law Theories: An Inquiry into Different Ways of Thinking* (OUP, 2017)

Biltgen P and Ryan S, *Activity-Based Intelligence: Principles and Applications* (Artech House, 2016)

Blacknell D and Griffiths H (eds), *Radar Automatic Target Recognition (ATR) and Non-Cooperative Target Recognition (NCTR)* (The Institute of Engineering and Technology, 2013)

Blix H, 'Means and Methods of Combat' in UNESCO, *International Dimensions of Humanitarian Law* (Martinus Nijhoff, 1988)

Boothby WH, *The Law of Targeting* (OUP, 2012)

—— 'How Far Will the Law Allow Unmanned Targeting to Go?' in Saxon D (ed), *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff, 2013)

—— *Conflict Law: The Influence of New Weapons Technology, Human Rights and Emerging Actors* (TMC Asser Press, 2014)

—— ‘Autonomous Attack – Opportunity or Spectre?’ in Gill TD (ed), *Yearbook of International Humanitarian Law 2013*, Vol. 16 (TMC Asser Press, 2015)

—— ‘Dehumanization: Is There a Legal Problem Under Article 36?’ in von Heinegg WH, Frau R and Singer T (eds), *Dehumanization of Warfare: Legal Implications of New Weapons Technologies* (Springer, 2018)

Bostrom N, *Superintelligence: Paths, Dangers, Strategies* (OUP, 2014)

Bubnicki Z, *Modern Control Theory* (Springer, 2005)

Cannizzaro E, ‘Proportionality in the Law of Armed Conflict’ in Clapham A and Gaeta P (eds), *The Oxford Handbook of International Law in Armed Conflict* (OUP, 2014)

Cawsey A, *The Essence of Artificial Intelligence* (Prentice Hall, 1998)

Chun WH and Papanikolopoulos N, ‘Robot Surveillance and Security’ in Siciliano B and Khatib O (eds), *Springer Handbook of Robotics* (Springer-Verlag, 2016)

Corn GS, ‘Autonomous Weapons Systems: Managing the Inevitability of ‘Taking the Man Out of the Loop’’ in Bhuta N et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (CUP, 2016)

Curtis J, ‘Relations Between Archaeologists and the Military in the Case of Iraq’ in Stone PG (ed), *Cultural Heritage, Ethics and the Military* (Boydell Press, 2011)

Darling K, ‘“Who’s Johnny?” Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy’ in Lin P, Abney K and Jenkins R (eds), *Robot Ethics 2.0* (OUP, 2017)

DeGrazia D, *Taking Animals Seriously: Mental Life and Moral Status* (CUP, 1996)

Dill J, *Legitimate Targets? Social Construction, International Law and US Bombing* (CUP, 2015)

Dinstein Y, *The Conduct of Hostilities Under the Law of International Armed Conflict* (3rd ed, CUP, 2016)

—— ‘Autonomous Weapons and International Humanitarian Law’ in von Heinegg WH, Frau R and Singer T (eds), *Dehumanization of Warfare: Legal Implications of New Weapons Technologies* (Springer, 2018)

Domingos P, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (Allen Lane, 2015)

Dwyer JG, *Moral Status and Human Life* (CUP, 2010)

Ekelhof MAC, 'Human Control in the Targeting Process' in Geiß R (ed), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016)

Flavell JH, 'Metacognitive Aspects of Problem Solving' in Resnick LB (ed), *The Nature of Intelligence* (Erlbaum Associates, 1976)

Frankish K and Ramsey WM (eds), *The Cambridge Handbook of Artificial Intelligence* (CUP, 2014)

Friedman B and Kahn PH, 'Human Agency and Responsible Computing: Implications for Computer System Design' in Friedman B (ed), *Human Values and the Design of Computer Technology* (CSLI, 1997)

Green LC, *The Contemporary Law of Armed Conflict* (2nd ed, MUP, 2000)

Hallevy G, *Liability for Crimes Involving Artificial Intelligence Systems* (Springer International, 2015)

Hambling D, *Swarm Troopers: How Small Drones Will Conquer the World* (Archangel Ink, 2015)

Henderson I, *The Contemporary Law of Targeting: Military Objectives, Proportionality and Precautions in Attack under Additional Protocol I* (Martinus Nijhoff, 2009)

—— and Bryan Cavanagh, 'Unmanned Aerial Vehicles (UAVs): Do They Pose Legal Challenges?' in Nasu H and McLaughlin R (eds), *New Technologies and the Law of Armed Conflict* (TMC Asser Press, 2014)

—— Keane P and Liddy J, 'Remote and Autonomous Warfare Systems: Precautions in Attack and Individual Accountability' in Ohlin J (ed), *Research Handbook on Remote Warfare* (Edward Elgar, 2017)

Hexmoor H et al, 'A Prospectus on Agent Autonomy' in Hexmoor H et al (eds), *Agent Autonomy* (Springer, 2003)

Homayounnejad M, 'Ensuring Fully Autonomous Weapons Systems Comply with the Rule of Distinction in Attack' in Casey-Maslen S et al (eds), *Drones and Other Unmanned Weapons Systems under International Law* (Brill Nijhoff, 2018)

Humphrey N, *Seeing Red: A Study in Consciousness* (Harvard University Press, 2006)

Jenks C, 'The Distraction of Full Autonomy and the Need to Refocus the CCW LAWS Discussion on Critical Functions' in Geiß R (ed), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016)

Kahneman D, *Thinking, Fast and Slow* (Penguin, 2012)

Kalmanovitz P, 'Judgment, Liability and the Risks of Riskless Warfare' in Bhuta N et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (CUP, 2016)

Kalshoven F, *Reflections on the Law of War: Collected Essays* (Martinus Nijhoff, 2007)

Kononenko I and Kukar M, *Machine Learning and Data Mining: Introduction to Principles and Algorithms* (Horwood, 2007)

Krishnan A, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate, 2009)

Langley P, *Elements of Machine Learning* (Morgan Kaufmann, 1996)

Lieblich E and Benvenisti E, 'The Obligation to Exercise Discretion in Warfare: Why Autonomous Weapons Systems are Unlawful' in Bhuta N et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (CUP, 2016)

Lin P, Bekey G and Abney K, 'Autonomous Military Robotics: Risks, Ethics, and Design', *US Department of Navy, Office of Naval Research* (2008) <http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil_fac> accessed 10 May 2018

Marauhn T, 'Meaningful Human Control – and the Politics of International Law' in von Heinegg WH, Frau R and Singer T (eds), *Dehumanization of Warfare: Legal Implications of New Weapons Technologies* (Springer, 2018)

Margulies P, 'Making Autonomous Targeting Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts' in Ohlin JD (ed), *Research Handbook on Remote Warfare* (Edward Elgar, 2017)

McCormack TLH and Durham H, 'Aerial Bombardment of Civilians: The Current International Legal Framework' in Tanaka Y and Young MB (eds), *Bombing Civilians: A Twentieth-Century History* (The New Press, 2009)

Metcalfe J and Son LK, 'Anoetic, Noetic, and Autonoetic Metacognition' in Beran MJ et al (eds), *Foundations of Metacognition* (OUP, 2012)

Newton M and May L, *Proportionality in International Law* (OUP, 2014)

Oeter S, 'Means and Methods of Combat' in Dieter Fleck (ed), *The Handbook of International Humanitarian Law* (3rd ed, OUP, 2013)

Pratzner PR, 'The Current Targeting Process', in Ducheine PAL, Schmitt MN and Osinga FPB (eds), *Targeting: The Challenges of Modern Warfare* (TMC Asser Press, 2016)

Proust J, *The Philosophy of Metacognition: Mental Agency and Self-Awareness* (OUP, 2013)

Quéguiner J-F, 'The Principle of Distinction: Beyond an Obligation of Customary International Humanitarian Law', in Hensel HM (ed), *The Legitimate Use of Military Force: The Just War Tradition and the Customary Law of Armed Conflict* (Routledge, 2016)

Richards NM and Smart WD, 'How Should the Law Think About Robots?' in Calo R, Froomkin AM and Kerr I (eds), *Robot Law* (Edward Elgar, 2016)

Rochlin GI, 'Iran Air Flight 655 and the USS *Vincennes*: Complex, Large-Scale Military Systems and the Failure of Control' in La Porte TR (ed), *Social Responses to Large Technical Systems: Control or Anticipation* (Springer Science + Business Media, 1991)

Roorda M, 'NATO's Targeting Process: Ensuring Human Control Over (and Lawful Use of) 'Autonomous' Weapons' in Williams AP and Scharre P (eds), *Autonomous Systems: Issues for Defence Policymakers* (Headquarters Supreme Allied Commander Transformation, 2015)

Russell S and Norvig P, *Artificial Intelligence: A Modern Approach* (3rd ed, Pearson, 2016)

Sammut C and Webb GI (eds), *Encyclopedia of Machine Learning* (Springer, 2010)

Sassóli M, Bouvier AA and Quintin A, *How Does the Law Protect in War?: Cases, Documents and Teaching Materials on Contemporary Practice in International Humanitarian Law, Vol. I* (3rd ed, ICRC, 2011)

Saxon D, 'What is 'Judgment' in the Context of the Design and Use of Autonomous Weapon Systems?' in Geiß R (ed), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016)

Schachter BJ, *Automatic Target Recognition* (3rd ed, SPIE Press, 2018)

Scharre P, *Army of None: Autonomous Weapons and the Future of War* (Norton, 2018)

Schmitt M, *Essays on Law and War at the Fault Lines* (TMC Asser Press, 2012)

Sharkey N, 'Staying in the Loop: Human Supervisory Control of Weapons' in Bhuta N et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (CUP, 2016)

Simon HA, *Models of Man: Social and Rational* (Wiley, 1957)

Simrock S, 'Control Theory' in Daniel Brandt (ed), *CAS CERN Accelerator School* (CERN, 2008) <https://cds.cern.ch/record/1100534/files/p73.pdf> accessed 8 May 2018

Singer P, *Practical Ethics* (2nd ed, CUP, 1993)

Singer PW, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (Penguin, 2009)

Smith BW, Lawyers and Engineers Should Speak the Same Robot Language' in Calo R, Froomkin AM and Kerr I (eds), *Robot Law* (Edward Elgar, 2016)

Solis GD, *The Law of Armed Conflict: International Humanitarian Law in War* (2nd ed, CUP, 2016)

Springer PJ, *Military Robots and Drones: A Reference Handbook* (ABC-CLIO, 2013)

Steinhardt RG, 'Weapons and the Human Rights Responsibilities of Multinational Corporations' in Stuart Casey-Maslen (ed), *Weapons Under International Human Rights Law* (CUP, 2014)

Suarez LB, *Control Theory Fundamentals* (Delve, 2017)

Suchman L, 'Situational Awareness and Adherence to the Principle of Distinction as a Necessary Condition for Lawful Autonomy' in Geiß R (ed), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016)

Taylor HR, *Data Acquisition for Sensor Systems* (Springer, 1997)

Thurnher JS, 'Examining Autonomous Weapon Systems from a Law of Armed Conflict Perspective' in Nasu H and McLaughlin R (eds), *New Technologies and the Law of Armed Conflict* (TMC Asser Press, 2014)

—— 'Means and Methods of the Future: Autonomous Systems' in Ducheine PAL, Schmitt MN and Osinga FPB (eds), *Targeting: The Challenges of Modern Warfare* (TMC Asser Press, 2016)

—— 'Feasible Precautions in Attack and Autonomous Weapons' in von Heinegg WH, Frau R and Singer T (eds), *Dehumanization of Warfare: Legal Implications of New Weapons Technologies* (Springer, 2018)

Trapp KN, 'A Framework of Analysis for Assessing Compliance of LAWS with IHL Precautionary Measures' in Geiß R (ed), *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security* (German Federal Foreign Office, 2016)

von Clausewitz C (author), Maude FN (ed), *On War*, Book I, Chapter III (Wordsworth Classics, 1997)

Wagner M, 'Autonomy in the Battlespace: Independently Operating Weapon Systems and the Law of Armed Conflict' in Saxon D (ed), *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff, 2013)

Walzer M, *Just and Unjust Wars* (5th ed, Basic Books, 2015)

Weatherall T, *Jus Cogens: International Law and Social Contract* (CUP, 2015)

White H, 'Civilian Immunity in the Precision-Guidance Age' in Primoratz I (ed), *Civilian Immunity in War* (OUP, 2010)

Wilmshurst E and Breau S (eds), *Perspectives on the ICRC Study on Customary International Humanitarian Law* (CUP, 2011)

Academic Articles

Albarella U, 'Archaeologists in Conflict: Empathizing with Which Victims?' (2009) 2 Heritage Management 105

Altmann J and Sauer F, 'Autonomous Weapon Systems and Strategic Stability' (2017) 59 Survival 117

Anderson K, Reisner D and Waxman MC, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems' (2014) 90 International Law Studies 386

Arkin RC, 'Lethal Autonomous Systems and the Plight of the Non-Combatant' (2013) 137 AISB Quarterly 4

Asaro P, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making' (2012) 94 International Review of the Red Cross 687

Aspray W, 'Back to Basics: The Stored Program Concept' (1990) 27 IEEE Spectrum 51

Atwood CP, 'Activity-Based Intelligence: Revolutionizing Military Intelligence Analysis' (2015) 77 Joint Force Quarterly 24

Babu GS and Suresh S, 'Meta-Cognitive Neural Network for Classification Problems in a Sequential Learning Framework' (2012) 81 Neurocomputing 86

Backstrom A and Henderson I, 'New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews' (2012) 94 International Review of the Red Cross 483

Benitez-Quiroz CF, Srinivasan R and Martinez AM, 'Facial Color is an Efficient Mechanism to Visually Transmit Emotion' (2018) 115 Proceedings of the National Academy of Sciences 3581 <http://www.pnas.org/content/pnas/115/14/3581.full.pdf> accessed 21 May 2018

Bishop M, 'Why Computers Can't Feel Pain' (2009) 19 Minds and Machine 519

Blank LR et al, 'Belligerent Targeting and the Invalidity of a Least Harmful Means Rule' (2013) 89 International Law Studies 536

Boon KE, ‘Are Control Tests Fit for the Future? The Slippage Problem in Attribution Doctrines’ (2014) 15 Melbourne Journal of International Law 329

Boothby WH, ‘Direct Participation in Hostilities – A Discussion of the ICRC Interpretive Guidance’ (2010) 1 International Humanitarian Legal Studies 143

Bostrom N, ‘The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents’ (2012) 22 Minds & Machines 71

Chengeta T, ‘Measuring Autonomous Weapon Systems Against International Humanitarian Law Rules’ (2016) 5 Journal of Law and Cyber Warfare 63

—— ‘Defining the Emerging Notion of ‘Meaningful Human Control’ in Weapon Systems’ (2017) 49 New York University Journal of International Law and Politics 833

Coglianese G and Lehr D, ‘Regulating by Robot: Administrative Decision Making in the Machine-Learning Era’ (2017) 105 Georgetown Law Journal 1147

Cohen A, ‘Rules and Standards in the Application of International Humanitarian Law’ (2008) 41 Israel Law Review 41

Corn GS, ‘Thinking the Unthinkable: Has the Time Come to Offer Combatant Immunity to Non-State Actors?’ (2011) 22 Stanford Law & Policy Review 253

—— ‘War, Law, and the Oft Overlooked Value of Process as a Precautionary Measure’ (2015) 42 Pepperdine Law Review 419

—— and Corn GP, ‘The Law of Operational Targeting: Viewing the LOAC Through an Operational Lens’ (2012) 47 Texas International Law Journal 337

—— and Schoettler JA, ‘Targeting and Civilian Risk Mitigation: The Essential Role of Precautionary Measures’ (2015) 223 Military Law Review 785

Cox MT, ‘Metacognition in Computation: A Selected Research Review’ (2005) 169 Artificial Intelligence 104

Crootof R, ‘The Killer Robots are Here: Legal and Policy Implications’ (2015) 36 Cardozo Law Review 1837

—— ‘A Meaningful Floor for “Meaningful Human Control”’ (2016) 30 Temple International & Comparative Law Journal 53

—— ‘War Torts: Accountability for Autonomous Weapons’ (2016) 164 University of Pennsylvania Law Review 1347

Denno DW, ‘A Mind to Blame: New Views on Involuntary Acts’ (2003) 21 Behavioural Sciences and the Law 601

Dickinson LA, 'Military Lawyers on the Battlefield: An Empirical Account of International Law Compliance' (2010) 104 American Journal of International Law 1

Dinstein Y, 'Legitimate Military Objectives under the Current Jus in Bello' (2002) 78 International Law Studies 139

Dunlap CJ Jr, 'Accountability and Autonomous Weapons: Much Ado About nothing?' (2016) 30 Temple International & Comparative Law Journal 63

Dyschkant A, 'Legal Personhood: How We Are Getting it Wrong' (2015) University of Illinois Law Review 2075

Egeland K, 'Lethal Autonomous Weapon Systems under International Humanitarian Law' (2016) 85 Nordic Journal of international Law 89

Ekelhof MAC, 'Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting' (2018) 71 Naval War College Review 61

Ernest N and Cohen K, 'Genetic Fuzzy Based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions' (2015) 6 Journal of Defense Management 139

Estreicher S, 'Privileging Asymmetric Warfare (Part II)?: The "Proportionality" Principle under International Humanitarian Law' (2011) 12 Chicago Journal of International Law 143

Farrant J and Ford CM, 'Autonomous Weapons and Weapons Reviews: The UK Second International Weapon Review Forum' (2017) 93 International Law Studies 389

Fenrick WJ, 'The Rule of Proportionality and Protocol I in Conventional Warfare' (1982) 98 Military Law Review 91

—— 'Targeting and Proportionality During the NATO Bombing Campaign Against Yugoslavia' (2001) 12 European Journal of International Law 489

Ford Lt Col CM, 'Autonomous Weapons and International Law' (2017) 69 South Carolina Law Review 413

Gardam JG, 'Proportionality and Force in International Law' (1993) 87 American Journal of International Law 391

Garner R and Alexander P, 'Metacognition: Answered and Unanswered Questions' (1989) 24 Educational Psychologist 143

Gillespie T, 'New Technologies and Design for the Laws of Armed Conflict' (2015) 160 The RUSI Journal 50

—— and West R, ‘Requirements for Autonomous Unmanned Air Systems Set by Legal Issues’ (2010) 4 The International C2 Journal 23

Goodman R, ‘The Power to Kill or Capture Enemy Combatants’ (2013) 24 The European Journal of International Law 819

Goodrich MA and Schultz AC, ‘Human-Robot Interaction: A Survey’ (2007) 1 Foundations and Trends in Human-Robot Interaction 203

Gracia D, ‘The Many Faces of Autonomy’ (2012) 33 Theoretical Medicine and Bioethics 57

Haulman DL, ‘The US Air Force in the Air War over Serbia, 1999’ (2015) 62 Air Power History 6

Horowitz MC, ‘Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems’ (2016) 30 Temple International & Comparative Law Journal 85

Jenks C, ‘The Gathering Swarm: The Path to Increasingly Autonomous Weapons Systems’ (2017) 57 Jurimetrics 341

Johns R and Davies GAM, ‘Civilian Casualties and Public Support for Military Action: Experimental Evidence’ (forthcoming) Journal of Conflict Resolution <https://doi.org/10.1177/0022002717729733> accessed 21 May 2018

Kaplow L, ‘Rules Versus Standards: An Economic Analysis’ (1992) 42 Duke Law Journal 557

Kuhn D and Dean D, ‘Metacognition: A Bridge Between Cognitive Psychology and Educational Practice’ (2004) 43 Theory into Practice 268

LeCun Y, Bengio Y and Hinton G, ‘Deep Learning’ (2015) 521 Nature Review 436

Lenat DB, ‘EURISKO: A Program That Learns New Heuristics and Domain Concepts’ (1983) 21 Artificial Intelligence 61

Lewis M, Sycara K and Scerri P, ‘Scaling Up Wide-Area-Search-Munition Teams’ (2009) 24 IEEE Intelligent Systems 10

Liu HY, ‘Categorization and Legality of Autonomous and Remote Weapons Systems’ (2012) 94 International Review of the Red Cross 627

Lostal M, Hausler K and Bongard P, ‘Armed Non-State Actors and Cultural Heritage in Armed Conflict’ (2017) 24 International Journal of Cultural Property 407

Manzotti R, ‘The Computational Stance is Unfit for Consciousness’ (2012) 4 International Journal of Machine Consciousness 401

Marchant GE et al, 'International Governance of Autonomous Military Robots' (2011) 12 Columbia Science and Technology Law Review 272

Marra WC and McNeil SK, 'Understanding 'The Loop': Regulating the Next Generation of War Machines' (2013) 36 Harvard Journal of Law & Public Policy 1139

Matambanadzo SM, 'The Body, Incorporated' (2013) 87 Tulane Law Review 1

McFarland T, 'Factors Shaping the Legal Implications of Increasingly Autonomous Military Systems', (2015) 900 International Review of the Red Cross 1313

McNeal GS, 'Targeted Killing and Accountability' (2014) 102 The Georgetown Law Journal 681

Meloni C, 'Command Responsibility: Mode of Liability for the Crimes of Subordinates or Separate Offence of the Superior?' (2007) 5 Journal of International Criminal Justice 619

Noorman M, 'Responsibility Practices and Unmanned Military Technologies' (2014) 20 Science and Engineering Ethics 809

Parks WH, 'Air War and the Law of War' (1990) 32 Air Force Law Review 1

Picker CB, 'A View from 40,000 Feet: International Law and the Invisible Hand of Technology' (2001) 23 Cardozo Law Review 149

Price WN, 'Black-Box Medicine' (2015) 28 Harvard Journal of Law & Technology 419

Quéguiner J-F, 'Precautions Under the Law Governing the Conduct of Hostilities' (2006) 88 International Review of the Red Cross 793

Ratches JA, 'Review of Current Aided/Automatic Target Acquisition Technology for Military Target Acquisition Tasks' (2011) 50 Optical Engineering 1

Ratner S and Slaughter A-M, 'Appraising Methods of International Law: A Prospectus for Readers' (1999) 93 American Journal of International Law 291

Reeves SR and Thurnher JS, 'Are We Reaching a Tipping Point? How Contemporary Challenges are Affecting the Military Necessity-Humanity Balance', *Harvard National Security Journal Features* (24 June 2013) <http://harvardnsj.org/wp-content/uploads/2013/06/HNSJ-Necessity-Humanity-Balance_PDF-format1.pdf> accessed 10 June 2018

Reggia JA, Monner D and Sylvester J, 'The Computational Explanatory Gap' (2014) 21 Journal of Consciousness Studies 153

Robertson HB Jr, 'The Principle of the Military Objective in the Law of Armed Conflict' (1997) 8 United States Air Force Academy Journal of Legal Studies 35

Rochat P, ‘Five Levels of Self-Awareness as They Unfold in Early Life’ (2003) 12 Consciousness and Cognition 717

Roff HM, ‘The Strategic Robot Problem: Lethal Autonomous Weapons in War’ (2014) 13 Journal of Military Ethics 211

Sassóli M, ‘Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified’ (2014) 90 International Law Studies 308

Savitha R, Suresh S and Kim HJ, ‘A Meta-Cognitive Learning Algorithm for an Extreme Learning Machine Classifier’ (2014) 6 Cognitive Computation 253

Scharre P, ‘Centaur Warfighting: The False Choice of Humans Vs. Automation’ (2016) 30 Temple International & Comparative Law Journal 151

Schlagel RH, ‘Why Not Artificial Consciousness or Thought?’ (1999) Minds and Machines 3

Schmitt MN, ‘The Principle of Distinction and Weapon Systems on the Contemporary Battlefield’ (2008) 7 Connections 46

—— ‘Human Shields in International Humanitarian Law’ (2009) 47 Columbia Journal of Transnational Law 292

—— ‘Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance’ (2010) 50 Virginia Journal of International Law 795

—— ‘The Interpretive Guidance on the Notion of Direct Participation in Hostilities’ (2010) 1 Harvard National Security Journal 5

—— ‘Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics’ (2013) Harvard National Security Journal Features

—— and Thurnher JS, ‘Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict’ (2013) 4 Harvard National Security Journal 231

—— and Widmar EW, ‘On Target: Precision and Balance in the Contemporary Law of Targeting’ (2014) 7 Journal of National Security Law & Policy 379

Schraw G, ‘Promoting General Metacognitive Awareness’ (1998) 26 Instructional Science 113

Searle J, ‘Minds, Brains, and Programs’ (1980) 3 Behavioral and Brain Sciences 417

Sharkey NE, ‘The Evitability of Autonomous Robot Warfare’ (2012) 94 International Review of the Red Cross 787

Silver D et al, ‘Mastering the Game of Go Without Human Knowledge’ (2017) 550 Nature 354

Smith BW, ‘Controlling Humans and Machines’ (2016) 30 Temple International & Comparative Law Journal 167

Solum LB, ‘Legal Personhood for the Artificial Intelligences’ (1992) 70 North Carolina Law Review 1231

Sparrow R, ‘Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender’ (2015) 91 International Law Studies 699

—— and Sparrow L, ‘In the Hands of Machines? The Future of Aged Care’ (2006) 16 Minds & Machines 141

Stein TL, ‘The Approach of the Different Drummer: The Principle of the Persistent Objector in International Law’ (1985) 26 Harvard International Law Journal 457

Stone P, ‘The Identification and Protection of Cultural Heritage During the Iraq Conflict: A Peculiarly English Tale’ (2005) 79 Antiquity 933

Talmon S, ‘International Law: The ICJ’s Methodology Between Induction, Deduction and Assertion’ (2015) 26 The European Journal of International Law 417

Thurnher JS, ‘No One at the Controls: Legal Implications of Fully Autonomous Targeting’ (2012) 67 Joint Force Quarterly 77

Turing AM, ‘Computing Machinery and Intelligence’ (1950) 59 Mind 433

van den Boogaard J, ‘Proportionality and Autonomous Weapons Systems’ (2015) 6 Journal of International Humanitarian Legal Studies 247

Wagner M, ‘Taking Humans Out of the Loop: Implications for International Humanitarian Law (2011) 21 Journal of Law, Information & Science 155

—— ‘The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems’ (2014) 47 Vanderbilt Journal of Transnational Law 1371

Watkin K, ‘Military Advantage: A Matter of Value, Strategy and Tactics’ (2014) 17 Yearbook of International Humanitarian Law 277

Watts S, ‘Autonomous Weapons: Regulation Tolerant or Regulation Resistant?’ (2016) 30 Temple International & Comparative Law Journal 177

Wright JD, ‘Excessive Ambiguity: Assessing and Refining the Proportionality Standard’ (2012) 94 International Review of the Red Cross 819

Legal Commentaries, Directives and Military Doctrine

Bothe M, Partsche KJ and Solf WA, *New Rules for Victims of Armed Conflict: Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949* (Martinus Nijhoff, 1982)

Dörmann K et al (eds), *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field* (CUP, 2016)

Doswald-Beck L (ed), *San Remo Manual on International Law Applicable to Armed Conflicts at Sea* (CUP, 1995)

Henckaerts JM and Doswald-Beck L, *Customary International Humanitarian Law, Vol. 1: Rules* (CUP, 2005) <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul> accessed 10 June 2018

International Committee of the Red Cross, *Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law* (ICRC, 2009)

Joint Chiefs of Staff, *Joint Publication 3-05.2: Doctrine for Joint Special Operations* (JCS, 21 May 2003)

—— *Joint Publication 3-60: Joint Targeting* (JCS, 31 January 2013)

—— *Joint Publication 1: Doctrine for the Armed Forces of the United States* (JCS, 25 March 2013, incorporating Change 1, 12 July 2017),

—— *DoD Dictionary of Military and Associated Terms* (JCS, August 2018) <http://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf> accessed 30 August 2018

Joint Staff, *Joint Doctrine Development Process*, CJCSM 5120.01A (Joint Staff, 29 December 2014)

Lee DH (ed), *Operational Law Handbook* (JAG's Legal Center & School, US Army, 2015)

NATO Military Committee, *NATO Rules of Engagement*, MC 362/1 (NATO HQ, 30 June 2003)

North Atlantic Treaty Organisation, *AJP-3.9: Allied Joint Doctrine for Joint Targeting* (Edition A Version 1, NATO Standardisation Office, April 2016)

North Atlantic Treaty Organisation, *AAP-47: Allied Joint Doctrine Development* (Edition B Version 1, NATO Standardisation Office, June 2016)

Nystuen G and Casey-Maslen S (eds), *The Convention on Cluster Munitions: A Commentary* (OUP, 2010)

Program on Humanitarian Policy & Conflict Research at Harvard University, *HPCR Manual on International Law Applicable to Air and Missile Warfare* (Harvard College, 2009)

—— *Commentary on the HPCR Manual on International Law Applicable to Air and Missile Warfare* (v2.1, Harvard College, 2010)

Sandoz Y, Swinarski C and Zimmermann B (eds), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Convention of 12th August 1949* (Martinus Nijhoff, 1987)

Schmitt MN (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd ed, CUP, 2017)

UK Ministry of Defence, *The Manual of the Law of Armed Conflict* (OUP, 2004)

—— *Joint Doctrine Publication 0-30.2: Unmanned Aircraft Systems* (Development, Concepts and Doctrine Centre, August 2017)

US Army, *Field Manual 3-60: The Targeting Process* (US Army Headquarters, November 2010)

—— *Army Doctrine Reference Publication 3-0: Unified Land Operations* (US Department of the Army, May 2012)

US Department of Defense, *Directive No. 5100.77: DoD Law of War Program* (9 December 1998) <https://biotech.law.lsu.edu/blaw/dodd/corres/pdf2/d510077p.pdf> accessed 22 September 2018

—— *Directive No. 3000.09: Autonomy in Weapon Systems* (21 November 2012, incorporating *Change 1*, 8 May 2017) <http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf> accessed 10 May 2018

—— *Law of War Manual* (DoD, 2015; December 2016 Update)

Reports, Working Papers and Conference Proceedings

AIV and CAVV 'Autonomous Weapon Systems: The Need for Meaningful Human Control' *Report No. 97 AIV/No. 26 CAVV* (AIV/CAVV, October 2015)

Alwardt, C and Krüger M, 'Autonomy of Weapon Systems, *IFSH/IFAR Food for Thought Paper* (February 2016) <https://ifsh.de/file-IFAR/pdf_english/IFAR_FFT_1_final.pdf> accessed 10 May 2018

Article 36, 'Structuring the Debate on Autonomous Weapons Systems', *Memorandum for Delegates to the Convention on Certain Conventional Weapons (CCW)* (14-15 November 2013) <http://www.article36.org/wp-content/uploads/2013/11/Autonomous-weapons-memo-for-CCW.pdf> accessed 10 June 2018

—— *Killing by Machine: Key Issues for Understanding Meaningful Human Control* (6 April 2015) <http://www.article36.org/wp-content/uploads/2013/06/KILLING_BY_MACHINE_6.4.15.pdf> accessed 10 May 2018

—— 'Key Elements of Meaningful Human Control', *Background Paper to Comments Prepared by Richard Moyes for the CCW Meeting of Experts on LAWS* (11-15 April 2016) <http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf> accessed 10 June 2018

—— 'Autonomous Weapon Systems : Evaluating the Capacity for 'Meaningful Human Control' in Weapon Review Processes', *Discussion Paper for the Group of Governmental Experts Meeting on LAWS* (13-17 November 2017) <http://www.article36.org/wp-content/uploads/2013/06/Evaluating-human-control-1.pdf> accessed 7 July 2018

Athalye A et al, 'Synthesizing Robust Adversarial Examples' (30 October 2017) <https://arxiv.org/pdf/1707.07397.pdf> accessed 16 May 2018

Behpour S, Kitani KM and Ziebart BD, 'ADA: A Game-Theoretic Perspective on Data Augmentation for Object Detection' (12 December 2017) <https://arxiv.org/pdf/1710.07735v2.pdf> accessed 13 May 2018

Bhargava P et al, 'The Robot Baby and Massive Metacognition: Future Vision', *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics*, IEEE *Xplore* (2012) 1 <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6400837> accessed 21 May 2018

Boulanin V and Verbruggen M, *Mapping the Development of Autonomy in Weapon Systems* (SIPRI, November 2017)

Brehm M, 'Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems Under International Humanitarian and Human Rights Law', *Geneva Academy Research Brief* (2017)

Brundage M et al, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (February 2018) <https://maliciousaireport.com/> accessed 16 May 2018

Buyers J, 'Liability Issues in Autonomous and Semi-Autonomous Systems', *Osborne Clarke Publication No. 0000000* (2015) <http://www.osborneclarke.com/media/filer_public/c9/73/c973bc5c-cef0-4e45-8554-f6f90f396256/itech_law.pdf> accessed 21 May 2018

Canning JS, 'A Concept of Operations for Armed Autonomous Systems', *Presentation at Third Annual Disruptive Technology Conference* (6-7 September 2006) <https://ndiastorage.blob.core.usgovcloudapi.net/ndia/2006/disruptive_tech/canning.pdf> accessed 21 August 2018

—— 'You've Just Been Disarmed. Have a Nice Day!' *IEEE Technology and Society Magazine* (Spring 2009) 12

Chrabaszcz P, Loshchilov I and Hutter F, 'Back to Basics: Benchmarking Canonical Evolution Strategies for Playing Atari' (24 February 2018) <https://arxiv.org/pdf/1802.08842.pdf> accessed 13 May 2018

Cummings ML, 'Creating Moral Buffers in Weapon Control Interface Design', *IEEE Technology and Society Magazine* (Fall 2004) 28

—— 'Artificial Intelligence and the Future of Warfare', *Chatham House Research Paper* (January 2017) <https://www.chathamhouse.org/publication/artificial-intelligence-and-future-warfare> accessed 10 May 2018

Dasgupta P, 'Distributed Automatic Target Recognition Using Multi-Agent UAV Swarms', *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (8-12 May 2006)

Davison N, 'A Legal Perspective: Autonomous Weapon Systems Under International Humanitarian Law' in UNODA, *Occasional Papers No. 30: Perspectives on Lethal Autonomous Weapon Systems* (United Nations, November 2017)

Defense Science Board, US Department of Defense, *The Role of Autonomy in DoD Systems* (Office of the Under Secretary of Defense for ATL, July 2012)

Dill J (Remarks by), 'Interpretive Complexity and the IHL Principle of Proportionality' (2014) 108 Proceedings of the Annual Meeting of the American Society of International Law 82

Gao P, Hensley R and Zielke A, 'A Road Map to the Future of the Auto Industry', *McKinsey Quarterly* (October 2014)

Garcia D, 'Governing Lethal Autonomous Weapon Systems', *Ethics & International Affairs* (13 December 2017) <https://www.ethicsandinternationalaffairs.org/2017/governing-lethal-autonomous-weapon-systems/> accessed 10 June 2018

Gaston EL, 'When Looks Could Kill: Emerging State Practice on Self-Defense and Hostile Intent', *Global Public Policy Institute Research Paper* (22 June 2017) <http://www.gppi.net/fileadmin/user_upload/media/pub/2017/gaston_2017_hostile-intent_web.pdf> accessed 7 July 2018

Geirhos R et al, 'Comparing Deep Neural Networks Against Humans: Object Recognition When the Signals Get Weaker' (21 June 2017) <https://arxiv.org/pdf/1706.06969.pdf> accessed 16 May 2018

Geiss R, *The International Law Dimension of Autonomous Weapons Systems* (Friedrich Ebert Stiftung Study, October 2015)

Goodfellow IJ, Shlens J and Szegedy C, 'Explaining and Harnessing Adversarial Examples' (v3, 20 March 2015) <https://arxiv.org/pdf/1412.6572.pdf> accessed 16 May 2018

Grace K et al, 'When Will AI Exceed Human Performance? Evidence from AI Experts' (v3, 20 May 2018) <https://arxiv.org/pdf/1705.08807.pdf> accessed 10 June 2018

Groves S, 'A Manual Adapting the Law of Armed Conflict to Lethal Autonomous Weapons Systems', *Margaret Thatcher Center for Freedom, Special Report No. 183*, (The Heritage Foundation, 7 April 2016) <http://thf-reports.s3.amazonaws.com/2016/SR183.pdf> accessed 4 October 2018

Gunning D, 'Explainable Artificial Intelligence (XAI)', *DARPA Program Information* (10 August 2016) <http://www.darpa.mil/program/explainable-artificial-intelligence> accessed 13 May 2018

—— 'Explainable Artificial Intelligence (XAI)', *DARPA Program Update* (November 2017) <https://www.darpa.mil/attachments/XAIProgramUpdate.pdf> accessed 13 May 2018

Hampson FJ (Remarks by), 'Proportionality and Necessity in the Gulf Conflict' (1992) 86 Proceedings of the Annual Meeting (American Society of International Law) 45

Hawley JK, 'Patriot Wars: Automation and the Patriot Air and Missile Defense System', *CNAS Ethical Autonomy Series* (January 2017) <https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf> accessed 10 May 2018

He K et al, 'Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNET Classification', *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)* (2015) 1026

—— 'Deep Residual Learning for Image Recognition' (10 December 2015) <https://arxiv.org/pdf/1512.03385.pdf> accessed 13 May 2018

Homayounnejad M, 'Regulating Lethal Autonomous Weapon Systems I: Assessing the Sense and Scope of 'Autonomy' in Emerging Military Weapon Systems', *TLI Think! Paper 76/2017* (2017) <https://ssrn.com/abstract=3027540> accessed 30 September 2018

—— 'Ensuring Lethal Autonomous Weapon Systems Comply with International Humanitarian Law, *TLI Think! Paper 85/2017* (2017) <https://ssrn.com/abstract=3073893> accessed 2 October 2018

—— 'Autonomous Weapon Systems, Drone Swarming and the Explosive Remnants of War', *TLI Think! Paper 1/2018* (2018) <https://ssrn.com/abstract=3099768> accessed 10 May 2018

—— 'The Lawful Use of Autonomous Weapon Systems for Targeted Strikes (Part 1): Concepts, Advantages and Technologies', *TLI Think! Paper 11/2018* (2018) <https://ssrn.com/abstract=3158170> accessed 4 October 2018

—— 'The Lawful Use of Autonomous Weapon Systems for Targeted Strikes (Part 2): Targeting Law & Practice', *TLI Think! Paper 13/2018* (2018) <https://ssrn.com/abstract=3200416> accessed 2 October 2018

—— and Overill RE, 'Preventing Autonomous Weapon Systems from Being Used to Perpetrate Intentional Violations of the Laws of War', *TLI Think! Paper 8/2018* (2018) <https://ssrn.com/abstract=3123254> accessed 21 May 2018

Horowitz MC et al, Strategic Competition in an Era of Artificial Intelligence', *CNAS Series on AI and International Security* (July 2018) <https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf?mtime=20180716122000> accessed 4 October 2018

Horowitz MC and Scharre P, 'Meaningful Human Control in Weapon Systems: A Primer', *CNAS Project on Ethical Autonomy Working Paper* (March 2015) <https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf> accessed 10 June 2018

Hsu J, 'Biggest Neural Network Ever Pushes AI Deep Learning', *IEEE Spectrum* (8 July 2015) <http://spectrum.ieee.org/tech-talk/computing/software/biggest-neural-network-ever-pushes-ai-deep-learning> accessed 13 May 2018

Human Rights Watch, *Off Target: The Conduct of the War and Civilian Casualties in Iraq* (Human Rights Watch, 2003)

—— *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch, 2012)

—— *Advancing the Debate on Killer Robots: 12 Key Arguments for a Preemptive Ban on Fully Autonomous Weapons* (Human Rights Watch, May 2014)

—— *Mind the Gap: The Lack of Accountability for Killer Robots* (Human Rights Watch, 2015)

—— *Precedent for Preemption: The Ban on Blinding Lasers as a Model for a Killer Robots Prohibition* (Human Rights Watch, November 2015)

—— *Making the Case: The Dangers of Killer Robots and the Need for a Preemptive Ban* (Human Rights Watch, 2016)

International Committee of the Red Cross, *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects* (Expert Meeting of 26-28 March 2014) (ICRC, November 2014)

—— 'Views of the International Committee of the Red Cross (ICRC) on Autonomous Weapon Systems', *Working Paper Submitted to the CCW Meeting of Experts on LAWS* (11-15 April 2016) <https://www.icrc.org/en/download/file/21606/ccw-autonomous-weapons-icrc-april-2016.pdf> accessed 21 May 2018

International Panel on the Regulation of Autonomous Weapons (iPRAW), 'Focus on Technology and Application of Autonomous Weapons, *"Focus on" Report No. 1* (August 2017)

—— 'Focus on Computational Methods in the Context of LAWS', *"Focus on" Report No. 2* (November 2017)

—— 'Focus on the Human-Machine Relations in LAWS', *"Focus on" Report No. 3* (March 2018)

JASON, *Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to the DoD* (JSR-16-Task-003, January 2017) <https://fas.org/irp/agency/dod/jason/ai-dod.pdf> accessed 16 May 2018

Jenzen-Jones NR (ed), *Indirect Fire: A Technical Analysis of the Employment, Accuracy, and Effects of Indirect-Fire Artillery Weapons* (ARES, January 2017) <https://www.icrc.org/en/document/indirect-fire-technical-analysis-employment-accuracy-and-effects-indirect-fire-artillery> accessed 17 August 2018

Joint Readiness Training Center, 'Operation OUTREACH: Tactics, Techniques, and Procedures', *News Letter No. 03-27* (October 2003)

Karpathy A and Fei-Fei L, 'Deep Visual-Semantic Alignments for Generating Image Description' (CVPR, 2015) <https://cs.stanford.edu/people/karpathy/cvpr2015.pdf> accessed 13 May 2018

Kereliuk C, Sturm BL and Larsen J, 'Deep Learning and Music Adversaries' (16 July 2015) <https://arxiv.org/pdf/1507.04761.pdf> accessed 16 May 2018

Krizhevsky A, Sutskever and Hinton GE, 'ImageNET Classification with Deep Convolutional Neural Networks, *Proceedings of the 25th International Conference on Neural Information Processing Systems* (2012) 1097

Lewis L, 'Redefining Human Control: Lessons From the Battlefield for Autonomous Weapons', *Center for Autonomy and AI* (March 2018) <https://www.cna.org/CNA_files/PDF/DOP-2018-U-017258-Final.pdf> accessed 10 June 2018

Murphy T VII, 'The First Level of Super Mario Bros. is Easy with Lexicographic Orderings and Time Travel…After That it Gets a Little Tricky' (1 April 2013) <https://www.cs.cmu.edu/~tom7/mario/mario.pdf> accessed 13 May 2018

Nguyen A, Yosinski J and Clune J, 'Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images', *Computer Vision and Pattern Recognition (CVPR '15)* (IEEE, 2015) <https://arxiv.org/pdf/1412.1897.pdf> accessed 16 May 2018

Office of the Prosecutor, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia* (2000) 39 ILM 1257

Office of the Secretary of Defense, *Unmanned Aircraft Systems Roadmap 2005-2030* (Department of Defense, 2005)

Offices of the Surgeon General, Multinational Force – Iraq and US Army Medical Command, *Mental Health Advisory Team* (*MHAT*) *IV: Operation Iraqi Freedom 05-07: Final Report* (17 November 2006) <http://www.combatreform.org/MHAT_IV_Report_17NOV06.pdf> accessed 21 May 2018

Roff H and Moyes R, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons', *Briefing Paper for Delegates at the CCW Meeting of Experts on LAWS* (11-15 April 2016) <http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf> accessed 7 July 2018

Rosert E, 'How to Regulate Autonomous Weapons: Steps to Codify Meaningful Human Control as a Principle of International Humanitarian Law', *PRIF Spotlight 6/2017* (November 2017) <https://www.hsfk.de/fileadmin/HSFK/hsfk_publikationen/Spotlight0617.pdf> accessed 10 June 2018

Sassóli M, 'Legitimate Targets of Attack Under International Humanitarian Law', *Harvard Program on Humanitarian Policy and Conflict Research, Background Paper* (2003)

Sauer F, 'ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting', *ICRAC News* (14 May 2014) <https://www.icrac.net/icrac-statement-on-technical-issues-to-the-2014-un-ccw-expert-meeting/> accessed 24 June 2018

Scharre P, *Robotics on the Battlefield Part II: The Coming Swarm* (CNAS, 2014) <https://s3.amazonaws.com/files.cnas.org/documents/CNAS_TheComingSwarm_Scharre.pdf> accessed 10 May 2018

—— 'Autonomous Weapons and Operational Risk', *CNAS Ethical Autonomy Project* (February 2016) <http://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf> accessed 22 September 2018

—— and Horowitz MC, 'An Introduction to Autonomy in Weapon Systems', *CNAS Ethical Autonomy Series Working Paper* (February 2015) <https://s3.amazonaws.com/files.cnas.org/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf> accessed 10 May 2018

Schmidhuber J, 'Deep Learning in Neural Networks: An Overview' (8 October 2014) <https://arxiv.org/pdf/1404.7828.pdf> accessed 13 May 2018

Sheridan TB and Verplank WL, *Human and Computer Control of Undersea Teleoperators* (Man-Machines Systems Laboratory, MIT, 1978)

Stillion J, *Trends in Air-to Air Combat: Implications for Future Air Superiority* (CSBA, 2015) <http://csbaonline.org/uploads/documents/Air-to-Air-Report-.pdf> accessed 17 August 2018

Su J, Vargas DV and Sakurai K, 'One Pixel Attack for Fooling Deep Neural Networks' (22 February 2018) <https://arxiv.org/pdf/1710.08864.pdf> accessed 16 May 2018

Szegedy C et al, 'Intriguing Properties of Neural Networks' (19 February 2014) <https://arxiv.org/pdf/1312.6199.pdf> accessed 16 May 2018

Ulgen O, 'Definition and Regulation of LAWS' *Submission to April 2018 GGE* (5 April 2018) <https://www.researchgate.net/publication/324227191_Dr_Ulgen_UN_GGE_LAWS_April_2018_-_submission_-_Definition_and_Regulation_of_LAWS> accessed 21 August 2018

United Nations Institute for Disarmament Research, 'The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward', *UNIDIR Resources, No. 2* (2014) <http://www.unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf> accessed 10 June 2018

—— 'The Weaponization of Increasingly Autonomous Technologies in the Maritime Environment: Testing the Waters', *UNIDIR Resources, No. 4* (2015) <http://www.unidir.org/files/publications/pdfs/testing-the-waters-en-634.pdf> accessed 10 May 2018

—— 'Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies', *UNIDIR Resources, No. 5* (2017) <http://www.unidir.org/files/publications/pdfs/safety-unintentional-risk-and-accidents-en-668.pdf> accessed 13 May 2018

—— 'The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches', *UNIDIR Resources, No. 6* (2017) <http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf> accessed 10 May 2018

US Air Force, *Unmanned Aircraft Systems Flight Plan, 2009-2047* (USAF HQ, 18 May 2009)

US Army, *Robotic and Autonomous Systems Strategy* (US Army Training & Doctrine Command, March 2017)

US Department of Defense, *Unmanned Systems Integrated Roadmap: FY 2013-2038* (DoD, 2014)

—— *Summary of the 2018 National Defense Strategy of the United States of America* (Department of Defense, January 2018)

Vanderelst D and Winfield A, 'The Dark Side of Ethical Robots' (*AIES 2018*, February 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_98.pdf> accessed 21 May 2018

Varshney KR and Alemzadeh H, 'On the Safety of Machine Learning: Cyber-Physical Systems, Decision Sciences, and Data Products' (v2 22 August 2017) <https://arxiv.org/pdf/1610.01256v2.pdf> accessed 7 July 2018

Watts BD, *Six Decades of Guided Munitions and Battle Networks: Progress and Prospects* (CSBA, March 2007) <http://csbaonline.org/publications/2007/03/six-decades-of-guided-munitions-and-battle-networks-progress-and-prospects/> accessed 3 July 2018

Weizmann N, 'Autonomous Weapon Systems under International Law', *Academy Briefing No. 8* (Geneva Academy of International Humanitarian Law and Human Rights, November 2014)

United Nations and Diplomatic Conference Documents

—— 'Chair's Summary of the Discussion on Agenda item 6(a) 9 and 10 April 2018, Agenda item 6(b) 11 April 2018 and 12 April 2018, Agenda item 6(c) 12 April 2018, Agenda item 6(d) 13 April 2018', *Chair's Documents at the 2018 GGE Meeting on LAWS* (9-13 April 2018) <https://www.unog.ch/80256EDD006B8954/(httpAssets)/DF486EE2B556C8A6C125827A00488B9E/$file/Summary+of+the+discussions+during+GGE+on+LAWS+April+2018.pdf> accessed 5 July 2018

—— France, Reservations and Declarations Made Upon Ratification of AP I (11 April 2001) <https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Notification.xsp?documentId=D8041036B40EBC44C1256A34004897B2&action=OpenDocument> accessed 4 October 2018

—— *Official Records of the Diplomatic Conference on the Reaffirmation and Development of International Humanitarian Law Applicable in Armed Conflicts*, Vol. VI (Federal Political Department, 1978)

—— *Report of the 2015 Informal Meeting of Experts on LAWS* (2 June 2015) UN Doc. CCW/MSP/2015/3 <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G15/111/60/PDF/G1511160.pdf?OpenElement> accessed 24 June 2018

—— *Report of the 2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems* (31 August 2018) UN Doc. CCW/GGE.2/2018/3 <https://www.unog.ch/80256EDD006B8954/(httpAssets)/20092911F6495FA7C125830E003F9A5B/$file/2018_GGE+LAWS_Final+Report.pdf> accessed 4 October 2018

—— UK, Reservations and Declarations Made Upon Ratification of AP I (28 January 1998) <https://ihl-databases.icrc.org/ihl/NORM/0A9E03F0F2EE757CC1256402003FB6D2?OpenDocument> accessed 4 October 2018

International Committee of the Red Cross, 'Statement of the International Committee of the Red Cross Under Agenda Item 6(b)', *ICRC Statement Delivered to the 2018 GGE on LAWS* (27-31 August 2018) <https://www.unog.ch/80256EDD006B8954/(httpAssets)/151EF67AD8224E14C125830600531382/$file/2018_GGE+LAWS+2_6b_ICRC.pdf> accessed 30 September 2018

UN Human Rights Council, *Report of the United Nations Fact-Finding Mission on the Gaza Conflict* (25 September 2009) UN Doc. A/HRC/12/48 <http://www2.ohchr.org/english/bodies/hrcouncil/docs/12session/A-HRC-12-48.pdf> accessed 13 October 2018

Blog Commentaries

Brandon H, 'Joint Series: Restricting Medical Personnel, Units, and Transports to 'Light Individual Weapons'', *Intercross Blog* (16 February 2017) <http://intercrossblog.icrc.org/blog/joint-series-restricting-medical-personnel-units-and-transports-to-light-individual-weapons> accessed 17 August 2018

Chengeta T, 'What Level of Human Control Over Autonomous Weapon Systems is Required by International Law?' *EJIL: Talk!* (17 May 2018) <https://www.ejiltalk.org/what-level-of-human-control-over-autonomous-weapon-systems-is-required-by-international-law/> accessed 10 June 2018

Christiano P, 'Approval-Directed Algorithmic Learning', *AI Alignment Blog* (21 February 2016) <https://ai-alignment.com/approval-directed-algorithm-learning-bf1f8fad42cd> accessed 7 July 2018

Crootof R, 'Why the Prohibition on Permanently Blinding Lasers is Poor Precedent for a Ban on Autonomous Weapon Systems', *Lawfare* (24 November 2015) <https://www.lawfareblog.com/why-prohibition-permanently-blinding-lasers-poor-precedent-ban-autonomous-weapon-systems> accessed 4 October 2018

Dunlap CJ Jr, 'To Ban New Weapons or Regulate Their Use?' *Just Security* (3 April 2015) <https://www.justsecurity.org/21766/guest-post-ban-weapons-regulate-use/> accessed 4 October 2018

Ekelhof M, 'Autonomous Weapons: Operationalizing Meaningful Human Control', *ICRC Humanitarian Law & Policy* (15 August 2018) <http://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/> accessed 21 August 2018

Gubrud MA, 'The Ottawa Definition of Landmines as a Start to Defining LAWS', *1.0 Human: Mark Gubrud's Weblog* (April 2018) <http://gubrud.net/wp-content/uploads/2018/04/Landmines_and_LAWS.pdf> accessed 7 July 2018

Homayounnejad M, 'Drone Swarming and the Explosive Remnants of War', *Opinio Juris* (19 March 2018) <http://opiniojuris.org/2018/03/19/drone-swarming-and-the-explosive-remnants-of-war/> accessed 10 May 2018

Lewis D, Modirzadeh N and Blum G, 'The Pentagon's New Algorithmic-Warfare Team', *Lawfare* (26 June 2017) <https://www.lawfareblog.com/pentagons-new-algorithmic-warfare-team> accessed 7 July 2018

Russell S, 'Of Myths and Moonshine', *Edge* (14 November 2014) <https://www.edge.org/conversation/the-myth-of-ai#26015> accessed 13 May 2018

Scharre P, 'Autonomy, "Killer Robots," and Human Control in the Use of Force – Part I', *Just Security* (9 July 2014) <https://www.justsecurity.org/12708/autonomy-killer-robots-human-control-force-part/> accessed 10 May 2018

Schmitt MN, 'Regulating Autonomous Weapons Might be Smarter than Banning Them', *Just Security* (10 August 2015) <https://www.justsecurity.org/25333/regulating-autonomous-weapons-smarter-banning/> accessed 13 May 2018

Solum L, 'Legal Theory Lexicon: Rules, Standards, and Principles', *Legal Theory Blog* (6 September 2009) <https://lsolum.typepad.com/legaltheory/2009/09/legal-theory-lexicon-rules-standards-and-principles.html> accessed 18 September 2018

Tan A, 'Responsibility and Control in International Law and Beyond', *The Hague Institute for Global Justice* (27 June 2013) <http://www.thehagueinstituteforglobaljustice.org/latest-insights/latest-insights/news-brief/responsibility-and-control-in-international-law-and-beyond/> accessed 10 June 2018

Other Sources*

—— 'Artificial Intelligence: Rise of the Machines', *The Economist Briefing* (9 May 2015)

—— 'Atlas: The World's Most Dynamic Humanoid', *Boston Dynamics* <http://www.bostondynamics.com/robot_Atlas.html> accessed 10 May 2018

* Press releases and press articles, encyclopaedic entries, letters, official memoranda, speeches, interviews, videos and websites.

—— ‘Bombing and Occupation of ICRC Facilities in Afghanistan’, *ICRC News Release 01/48* (26 October 2001) <https://www.icrc.org/eng/resources/documents/news-release/2009-and-earlier/57jrdx.htm> accessed 21 August 2018

—— ‘Boomerang III: State-of-the-Art Shooter Detection’, *Raytheon* <https://www.raytheon.com/capabilities/products/boomerang> accessed 18 May 2018

—— ‘Boomerang Warrior-X: Wearable Shooter Detection System for Soldiers’, *Raytheon* <https://www.raytheon.com/capabilities/products/boomerang_warriorx> accessed 18 May 2018

—— ‘Camopedia: The Camouflage Encyclopedia’ <http://camopedia.org/index.php?title=Main_Page> accessed 11 August 2018

—— ‘CBU-105 Sensor Fuzed Weapon: USAF’s Ultimate Tank-Buster’, *DefenCyclopedia* (12 June 2015) <https://defencyclopedia.com/2015/06/12/cbu-105-sensor-fuzed-weapon-usafs-ultimate-tank-buster/> accessed 7 July 2018

—— ‘Comments of the National PNT Advisory Board’, *Jamming the Global Positioning System – A National Security Threat: Recent Events and Potential Cures* (4 November 2010) <http://www.gps.gov/governance/advisory/recommendations/2010-11-jammingwhitepaper.pdf> accessed 18 May 2018

—— ‘Coyote UAS’, *Raytheon* <http://www.raytheon.com/capabilities/products/coyote/> accessed 10 May 2018

—— ‘Explained: How Cruise Missiles Work’, *DefenCyclopedia* (1 August 2014) <https://defencyclopedia.com/2014/08/01/explained-how-cruise-missiles-work/> accessed 7 July 2018

—— ‘GPS: The Global Positioning System’ <http://www.gps.gov/> accessed 18 May 2018

—— ‘Intelligence Technology to Keep Joint Force Command One Step Ahead of Adversaries’, *Ministry of Defence News* (17 July 2018) <https://www.gov.uk/government/news/intelligence-technology-to-keep-joint-force-command-one-step-ahead-of-adversaries> Accessed 18 July 2018

—— ‘ICRC Warehouses Bombed in Kabul’, *ICRC News Release 01/43* (16 October 2001) <https://www.icrc.org/eng/resources/documents/news-release/2009-and-earlier/57jrcz.htm> accessed 21 August 2018

—— ‘Iran Shows ‘Hacked US Spy Drone’ Video Footage’, *BBC News* (7 February 2013) <http://www.bbc.co.uk/news/world-middle-east-21373353> accessed 18 May 2018

—— 'Kunduz Bombing: US Attacked MSF Clinic 'In Error'', *BBC News* (25 November 2015) <http://www.bbc.co.uk/news/world-asia-34925237> accessed 21 August 2018

—— 'List of World Heritage in Danger' <https://whc.unesco.org/en/danger/> accessed 21 August 2018

—— 'Memorandum for Assistant General Counsel (International), Office of the Secretary of Defense, 1977 Protocols Additional to the Geneva Conventions: Customary International Law Implications' (8 May 1986)

—— *Nature Reviews* Glossary <http://www.nature.com/nrg/journal/v10/n6/glossary/nrg2579.html> accessed 1 June 2016

—— 'Pride and Prejudice: The Odds on a Conflict Between the Great Powers', *The Economist Special Report: The Future of War* (27 January 2018)

—— 'Rockwell Collins Delivers Latest Digital GPS Receiver Technology to US Air Force Special Operations Command', *Rockwell Collins News* (24 August 2017) <https://www.rockwellcollins.com/Data/News/2017-Cal-Yr/GS/20170824-DIGAR-delivery.aspx> accessed 8 September 2018

—— 'Rockwell Collins Successfully Demonstrates 5,000 mile Next-Generation Wideband High Frequency Communications Link', *Rockwell Collins News* (18 September 2017) <https://www.rockwellcollins.com/Data/News/2017-Cal-Yr/GS/20170918-PACAF-HF-Comms.aspx> accessed 8 September 2018

—— 'Samsung Techwin SGR-A1 Sentry Guard Robot', *GlobalSecurity.org* (7 November 2011) <http://www.globalsecurity.org/military/world/rok/sgr-a1.htm> accessed 17 August 2018

—— 'Smart Weapons: The Vision Thing', *The Economist* (3 December 2016)

—— 'The Newest Thing in Quantum Imaging', *DoD Armed with Science* (3 January 2014) <http://science.dodlive.mil/2014/01/03/the-newest-thing-in-quantum-imaging/> accessed 14 May 2018

—— 'Wideband High Frequency Communications Provide Net-Centric, High-Speed Beyond Line of Sight Communications in Anti-Access Area-Denial (A2/AD) Battlefield Environments', *International Defence, Security & Technology* (11 August 2017) <http://idstch.com/home5/international-defence-security-and-technology/technology/ict/wideband-high-frequency-communications-provide-net-centric-high-speed-beyond-line-sight-communications-anti-accessarea-denial-a2ad-battlefield-environments/> accessed 8 September 2018

—— 'World Heritage List' <https://whc.unesco.org/en/list/> accessed 21 August 2018

Allison I, 'When Intelligent Algorithms Start Spoofing Each Other, Regulation Becomes a Science', *International Business Times* (29 June 2016) <http://www.ibtimes.co.uk/machine-learning-markets-when-intelligent-algorithms-start-spoofing-each-other-regulation-becomes-1567986> accessed 13 May 2018

Associated Press, 'Putin: Leader in Artificial Intelligence Will Rule World', *CNBC* (4 September 2017) <https://www.cnbc.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html> accessed 4 October 2018

Bailey K, 'Pseudolites Preserve Position Information During GPS-Denied Conditions', *US Army Press Release* (2 June 2016) <https://www.army.mil/article/169033/pseudolites_preserve_position_information_during_gps_denied_conditions> accessed 8 September 2018

Beckhausen R, 'Chinese Robo-Boats Swarm the South China Sea', *War is Boring* (31 May 2018) <https://warisboring.com/dozens-of-chinese-robot-boats-swarm-the-sea/> accessed 3 June 2018

Bellinger III J and Haynes WJ, 'Letter from John Bellinger III, Legal Adviser, US Dept. of State, and William J. Haynes, General Counsel, US Dept. of Defense, to Dr Jakob Kellenberger, President, International Committee of the Red Cross, Regarding Customary International Law Study [3 November 2006] (2007) 46 ILM 514

Boyd JR, 'Patterns of Conflict' in Chet Richards and Chuck Spinney (eds), *Defense and the National Interest* (January 2007) <http://www.dnipogo.org/boyd/patterns_ppt.pdf> accessed 8 May 2018

Campaign to Stop Killer Robots, *Campaign to Stop Killer Robots Website* <https://www.stopkillerrobots.org/ > accessed 10 May 2018

—— *Retaining Human Control of Weapons Systems* (9-13 April 2018) <https://www.stopkillerrobots.org/wp-content/uploads/2018/03/KRC_Briefing_CCWApr2018.pdf> accessed 4 October 2018

Carroll G, 'On Target with Boomerang III: Acoustic Sensing Technology', *Army Technology* (30 May 2012) <https://www.army-technology.com/features/featuredssi-boomerang-iii-dstl/> accessed 18 May 2018

Carter General Sir N, 'Dynamic Security Threats and the British Army', *Speech Delivered to the Royal United Services Institute* (22 January 2018) <https://rusi.org/event/dynamic-security-threats-and-british-army> accessed 4 October 2018

Clark C, The Terminator Conundrum: VCJCS Selva on Thinking Weapons', *Breaking Defense* (21 January 2016) <https://breakingdefense.com/2016/01/the-terminator-conundrum-vcjcs-selva-on-thinking-weapons/> accessed 4 October 2018

Coggins KM, 'Position, Navigation and Timing and What it Means for the Soldier', *DoD Armed with Science* (27 February 2016) <http://science.dodlive.mil/2016/02/27/staying-on-course-positioning-navigation-and-timing-pnt-and-what-it-means-for-the-soldier/> accessed 18 May 2018

Corrigan J, 'Three-Star General Wants AI in Every New Weapon System', *Defense One* (3 November 2017) <https://www.defenseone.com/technology/2017/11/three-star-general-wants-artificial-intelligence-every-new-weapon-system/142239/?oref=defenseone_today_nl> accessed 7 July 2018

Cox M, 'Army Eyes Autonomous Convoys to Prevent Future Casualties', *Military.com* (1 May 2018) <https://www.military.com/dodbuzz/2018/05/01/army-eyes-autonomous-convoys-prevent-future-casualties.html> accessed 10 May 2018

DARPA, 'Justification Book Vol. 1: Research, Development, Test & Evaluation, Defense-Wide', *Department of Defense FY 2013 President's Budget Submission* <https://www.darpa.mil/attachments/(2G4)%20Global%20Nav%20-%20About%20Us%20-%20Budget%20-%20Budget%20Entries%20-%20FY2013%20(Approved).pdf> accessed 10 May 2018

—— 'Broad Agency Announcement: Target Recognition and Adaptation in Contested Environments (TRACE)', *Strategic Technology Office: DARPA BAA-15-09* (1 December 2014) <https://www.fbo.gov/utils/view?id=d68e7ccabd593f2f46b2328fa18dc6db> accessed 14 May 2018

—— 'Broad Agency Announcement: Collaborative Operations in Denied Environment (CODE)', *Strategic Technology Office: DARPA BAA-14-33* (25 April 2014) <https://www.fbo.gov/utils/view?id=260d113392090305f63b281cf20bfea9> accessed 10 May 2018

—— 'ACTUV "Sea Hunter" Prototype Transitions to Office of Naval Research for Further Development', *DARPA News and Events* (30 January 2018) <https://www.darpa.mil/news-events/2018-01-30a> accessed 10 May 2018

Davey T, 'How AI Handles Uncertainty: An Interview with Brian Zeibart', *Future of Life Institute News* (15 March 2018) <https://futureoflife.org/2018/03/15/how-ai-handles-uncertainty-brian-ziebart/> accessed 13 May 2018

Deputy Secretary of Defense Memorandum, 'Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven)' (26 April 2017) <https://www.govexec.com/media/gbc/docs/pdfs_edit/establishment_of_the_awcft_project_maven.pdf> accessed 7 July 2018

Dombe AR, 'Biometric Target Recognition', *Israel Defense* (16 March 2017) <https://www.israeldefense.co.il/en/node/28881> accessed 7 July

Etzioni A and Etzioni O, 'Pros and Cons of Autonomous Weapons Systems', *Military Review* (May-June 2017)

Future of Life Institute, 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers' (28 July 2015) <https://futureoflife.org/open-letter-autonomous-weapons/> accessed 4 October 2018

—— 'An Open Letter to the United Nations Convention on Certain Conventional Weapons' (21 August 2017) <https://futureoflife.org/autonomous-weapons-open-letter-2017> accessed 4 October 2018

Gibbs S, 'Chatbot Lawyer Overturns 160,000 Parking Tickets in London and New York', *The Guardian* (28 June 2016) <https://www.theguardian.com/technology/2016/jun/28/chatbot-ai-lawyer-donotpay-parking-tickets-london-new-york> accessed 21 May 2018

Goodfellow I, *Presentation at Re-Work Deep Learning Summit* (24 February 2015) <https://www.youtube.com/watch?v=Pq4A2mPCB0Y> accessed 16 May 2018

Gorman J, 'Target Recognition and Adaptation in Contested Environments (TRACE)', *DARPA Program Information* <https://www.darpa.mil/program/trace> accessed 14 May 2018

Hambling D, 'US Navy Plans to Fly First Drone Swarm this Summer', *DefenseTech* (4 January 2016) <https://www.defensetech.org/2016/01/04/u-s-navy-plans-to-fly-first-drone-swarm-this-summer/> accessed 10 May 2018

Harding L and Engel M, 'US Bomb Blunder Kills 30 at Afghan Wedding', *The Guardian* (2 July 2002) <https://www.theguardian.com/world/2002/jul/02/afghanistan.lukeharding> accessed 21 August 2018

Hecht J, 'Did Iran Capture US Drone by Hacking its GPS Signal?' *New Scientist* (16 December 2011)

Horowitz MC and Scharre P, 'Do Killer Robots Save Lives?' *Politico Magazine* (19 November 2014) <http://www.politico.com/magazine/story/2014/11/killer-robots-save-lives-113010> accessed 3 July 2018

Human Rights Watch, 'Ukraine: Unguided Rockets Killing Civilians', *Human Rights Watch News* (24 July 2014) <https://www.hrw.org/news/2014/07/24/ukraine-unguided-rockets-killing-civilians> accessed 3 July 2018

International Committee for Robot Arms Control, 'Statements: Mission Statement and Berlin Statement' <https://www.icrac.net/statements/> accessed 4 October 2018

International Committee of the Red Cross, 'Autonomous Weapon Systems – Q&A', *ICRC Article* (12 November 2014) <https://www.icrc.org/en/document/autonomous-weapon-systems-challenge-human-control-over-use-force> accessed 10 June 2018

Israeli Defense Forces, 'How is the IDF Minimizing Harm to Civilians in Gaza?' *IDF Blog* (16 July 2014) <https://www.idf.il/en/minisites/hamas/how-is-the-idf-minimizing-harm-to-civilians-in-gaza/> accessed 8 September 2018

Jackson Colonel R, 'Autonomous Weaponry and Armed Conflict', *ASIL Panel Discussion* (10 April 2014) <https://www.youtube.com/watch?v=duq3DtFJtWg> accessed 10 May 2018

Jaworska A and Tannenbaum J, 'The Grounds of Moral Status' in Zalta EN (ed) *Stanford Encyclopedia of Philosophy* (Spring 2018) <https://plato.stanford.edu/archives/spr2018/entries/grounds-moral-status/> accessed 21 May 2018

Jones M, 'Army Pseudolites: What, Why and How?' *GPS World* (9 August 2017) <http://gpsworld.com/army-pseudolites-what-why-and-how/> accessed 8 September 2018

Kania EB, 'China is On a Whole-of-Nation Push for AI. The US Must Match It', *Defense One* (8 December 2017) <https://www.defenseone.com/ideas/2017/12/us-china-artificial-intelligence/144414/> accessed 4 October 2018

Kellenberger J, *Strengthening Legal Protection for Victims of Armed Conflicts: States' Consultations and Way Forward* (ICRC, 12 May 2011) <https://www.icrc.org/en/doc/assets/files/red-cross-crescent-movement/31st-international-conference/icrc-president-statement-2010-05-12.pdf> accessed 26 March 2019

Keller J, 'DARPA TRACE Program Using Advanced Algorithms, Embedded Computing for Radar Target Recognition', *Military and Aerospace Electronics* (24 July 2015) <http://www.militaryaerospace.com/articles/2015/07/hpec-radar-target-recognition.html> accessed 14 May 2018

Kelly K, 'The Three Breakthroughs That Have Finally Unleashed AI on the World', *Wired* (27 October 2014) <http://www.wired.com/2014/10/future-of-artificial-intelligence/> accessed 4 October 2018

Komar R, 'How to Digitally Verify Combatant Affiliation in Middle East Conflicts', *Bellingcat* (9 July 2018) <https://www.bellingcat.com/resources/how-tos/2018/07/09/digitally-verify-middle-east-conflicts/> accessed 11 August 2018

Ledé JC, 'Collaborative Operations in Denied Environment (CODE)', *DARPA Program Information*, <https://www.darpa.mil/program/collaborative-operations-in-denied-environment> accessed 10 May 2018

Lloyd-Jones Lord, 'General Principles of Law in International Law and Common Law', *Speech Delivered to the Conseil d'Etat, Paris* (16 February 2018) <https://www.supremecourt.uk/docs/speech-180216.pdf> Accessed 1 April 2019

Manea O, 'The Role of Offset Strategies in Restoring Conventional Deterrence' (Interview with US Deputy Secretary of Defense, Robert O. Work), *Small Wars Journal* (4 January 2018) http://smallwarsjournal.com/jrnl/art/role-offset-strategies-restoring-conventional-deterrence> accessed 4 October 2018

Mehta A, 'DoD Weapons Designer: Swarming Teams of Drones Will Dominate Future Wars', *Defense News* (30 March 2017) <https://www.defensenews.com/smr/unmanned-unleashed/2017/03/30/dod-weapons-designer-swarming-teams-of-drones-will-dominate-future-wars/> accessed 10 May 2018

Metz C, 'Inside Libratus, The Poker AI That Out-Bluffed the Best Humans, *Wired* (1 February 2017) <https://www.wired.com/2017/02/libratus/> accessed 4 October 2018

Moyes R, 'Autonomous Weapons Systems Policy', *Talk Delivered for MIT Course 6.S099: Artificial General Intelligence* (17 April 2018) <https://www.youtube.com/watch?v=U6lJI-NSfBY&t=2847s> accessed 8 May 2018

O'Connor JM, 'Applying the Law of Targeting to the Modern Battlefield', *Speech Delivered by DoD General Counsel to New York University School of Law* (28 November 2016) <https://www.defense.gov/Portals/1/Documents/pubs/Applying-the-Law-of-Targeting-to-the-Modern-Battlefield.pdf> accessed 7 July 2018

Osborne K, 'Air Force Seeks Swarms of Attack Mini-Drones' (Interview with Air Force Chief Scientist Gregory Zacharias), *Scout Warrior* (10 May 2016) <https://www.wearethemighty.com/articles/air-force-seeks-swarms-of-versatile-mini-drones> accessed 10 May 2018

Pellerin C, 'Deputy Secretary: Third Offset Strategy Bolsters America's Military Deterrence', *US Department of Defense News* (31 October 2016) <https://www.defense.gov/News/Article/Article/991434/deputy-secretary-third-offset-strategy-bolsters-americas-military-deterrence/> accessed 30 September 2018

Pentagon Memorandum, 'The Defense Innovation Initiative' (15 November 2014) <http://archive.defense.gov/pubs/OSD013411-14.pdf> accessed 30 September 2018

Reagan R, 'Letter of Transmittal to the United States Senate' (29 January 1987), reprinted in (1987) 81 American Journal of International Law 910

Richter W, 'Military Rationale for Autonomous Functions in Weapons Systems', *Presentation at the 2015 Meeting of Experts on LAWS* (13-17 April 2015)

Romeny BH, 'Tutorial: Deep Learning in Human and Computer Vision' (30 August 2017) <http://bmia.bmt.tue.nl/people/BRomeny/Courses/Taipei2017/index.html> accessed 13 May 2018

Schmitt MN, 'Autonomous Weapons Systems and International Law', *LENS Conference 2016: Autonomous Weapons in the Age of Hybrid War* (27 February 2016) <https://www.youtube.com/watch?v=b5mz7Y2FmU4> accessed 11 August 2018

Sherwood H, 'Israel Using Flechette Shells in Gaza', *The Guardian* (20 July 2014) <https://www.theguardian.com/world/2014/jul/20/israel-using-flechette-shells-in-gaza> accessed 13 October 2018

Smalley D, 'The Future is Now: Navy's Autonomous Swarm Boats can Overwhelm Adversaries', *Office of Naval Research News & Media Center* (5 October 2014) https://www.onr.navy.mil/en/Media-Center/Press-Releases/2014/autonomous-swarm-boat-unmanned-caracas> accessed 10 May 2018

Tamkin E and McLeary P, 'China Seizes US Navy Drone in South China Sea', *Foreign Policy* (16 December 2016) <https://foreignpolicy.com/2016/12/16/china-seizes-u-s-navy-drone-in-in-south-china-sea-raising-stakes-president-trump/> accessed 7 July 2018

Thomas D, 'Microsoft Pulls Twitter Bot Tay after Racist Tweets', *Financial Times* (24 March 2016) <https://www.ft.com/content/8ba60bc4-f1c0-11e5-aff5-19b4e253664a> accessed 13 May 2018

Thurnher J and Kelly T, 'Panel Discussion: Collateral Damage Estimation', *US Naval War College* (23 October 2012) <https://www.youtube.com/watch?v=AvdXJV-N56A> accessed 6 September 2018

Tucker P, 'The Future the US Military is Constructing: a Giant, Armed Nervous System', *Defense One* (26 September 2017) <http://www.defenseone.com/technology/2017/09/future-us-military-constructing-giant-armed-nervous-system/141303/> accessed 8 September 2018

—— 'Russia to the United Nations: Don't Try to Stop Us From Building Killer Robots', *Defense One* (21 November 2017) <https://www.defenseone.com/technology/2017/11/russia-united-nations-dont-try-stop-us-building-killer-robots/142734/> accessed 4 October 2018

—— 'The US Navy is Developing Mothership Drones for Coastal Defense', *Defense One* (1 June 2018) <https://www.defenseone.com/technology/2018/06/us-navy-developing-mothership-drones-coastal-defense/148671/?oref=d-dontmiss> accessed 13 June 2018

US Department of Defense 'Department of Defense Announces Successful Micro-Drone Demonstration', *Press Release No. NR-008-17* (9 January 2017) <https://www.defense.gov/News/News-Releases/News-Release-View/Article/1044811/department-of-defense-announces-successful-micro-drone-demonstration/> accessed 10 May 2018

—— 'Joint Enterprise Defense Infrastructure (JEDI)', *US Department of Defense Memo* (6 November 2017) <https://www.nextgov.com/media/gbc/docs/pdfs_edit/121217fk1ng.pdf> accessed 8 September 2018

Vinson B, 'X-47B Makes First Arrested Landing at Sea', *Navy News Services* (10 July 2013) <http://www.navy.mil/submit/display.asp?story_id=75298> accessed 10 May 2018

Wagner M, 'Autonomous Weapon Systems', *Max Planck Encyclopedia of Public International Law* (2016) <http://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e2134> accessed 4 October 2018

Weisgerber M, 'The Pentagon's New Algorithmic Warfare Cell Gets Its First Mission: Hunt ISIS', *Defense One* (14 May 2017) <http://www.defenseone.com/technology/2017/05/pentagons-new-algorithmic-warfare-cell-gets-its-first-mission-hunt-isis/137833/> accessed 7 July 2018

—— 'The Pentagon's New Artificial Intelligence is Already Hunting Terrorists, *Defense One* (21 December 2017) <http://www.defenseone.com/technology/2017/12/pentagons-new-artificial-intelligence-already-hunting-terrorists/144742/> accessed 7 July 2018

Work RO, 'Deputy Secretary of Defense Speech', *CNAS Defense Forum* (14 December 2015) <http://www.defense.gov/News/Speeches/Speech-View/Article/634214/cnas-defense-forum> accessed 4 October 2018